UNIVERSIDADE FEDERAL DE PELOTAS
Centro de Desenvolvimento Tecnológico
Programa de Pós-Graduação em Computação

Tese

Um estudo sobre aprendizado de máquina multimodal para identificar a depressão em adolescentes brasileiros usando dados coletados via *smartphones*

Ramásio Ferreira de Melo

Pelotas, 2023

Ramásio Ferreira de Melo

Um estudo sobre aprendizado de máquina multimodal para identificar a depressão em adolescentes brasileiros usando dados coletados via *smartphones*

Tese apresentada ao Programa de Pós-Graduação em Computação do Centro de Desenvolvimento Tecnológico da Universidade Federal de Pelotas, como requisito parcial à obtenção do título de Doutor em Ciência da Computação.

Orientador: Prof. Dr. Ricardo Matsumura Araújo
Coorientador: Prof. Dr. Christian Costa Kieling

Pelotas, 2023

Universidade Federal de Pelotas / Sistema de Bibliotecas
Catalogação na Publicação

M528e Melo, Ramásio Ferreira de

Um estudo sobre aprendizado de máquina multimodal para identificar a depressão em adolescentes brasileiros usando dados coletados via smartphones / Ramásio Ferreira de Melo ; Ricardo Matsumura Araújo, orientador ; Christian Costa Kieling, coorientador. — Pelotas, 2023.

179 f. : il.

Tese (Doutorado) — Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, 2023.

1. Depressão. 2. Saúde mental. 3. Inteligência artificial. 4. Sensoriamento passivo. 5. Smartphone. I. Araújo, Ricardo Matsumura, orient. II. Kieling, Christian Costa, coorient. III. Título.

CDD : 005

Elaborada por Simone Godinho Maisonave CRB: 10/1733

Ramásio Ferreira de Melo

Um estudo sobre aprendizado de máquina multimodal para identificar a depressão em adolescentes brasileiros usando dados coletados via *smartphones*

Tese aprovada, como requisito parcial, para obtenção do grau de Doutor em Ciência da Computação, Programa de Pós-Graduação em Computação, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas.

Data da Defesa: 25 de agosto de 2023

Banca Examinadora:

Prof. Dr. Ricardo Matsumura Araújo (Orientador)
Doutor em Computação pela Universidade Federal do Rio Grande do Sul.

Prof. Dr. Ives Passos
Doutor em Psiquiatria pela Universidade Federal do Rio Grande do Sul.

Prof. Dr. Ulisses Brisolara Corrêa
Doutor em Computação pela Universidade Federal de Pelotas.

Prof. Dr. Bruno Pereira Nunes
Doutor em Epidemiologia pela Universidade Federal de Pelotas.

Dedico este trabalho primeiramente a DEUS, a meus pais, Adauto e Maria Onélia, e minha esposa Priscila, e especialmente à minha filha, Cecília, que nasceu durante o período do doutorado. Você é minha maior motivação para alcançar meus objetivos. Agradeço por estarem presentes em todos os momentos.

AGRADECIMENTOS

Em primeiro lugar, expresso minha profunda gratidão a Deus por me proteger, guiar e fortalecer durante todo o processo. Sem Sua presença e orientação, essa conquista não seria possível.

Ao Professor Christian e aos colegas do PRODIA, sou grato por sua orientação e contribuição valiosa de todos para o desenvolvimento desta tese e pela visão mais ampla da humanidade proporcionada.

Ao Professor Ricardo, agradeço imensamente pela oportunidade, pelos ensinamentos compartilhados e pelo exemplo de dedicação. Sua contribuição foi fundamental para materializar este trabalho.

À minha amada esposa, Priscila, sou imensamente grato pelo companheirismo e apoio nos momentos difíceis ao longo dessa jornada acadêmica.

Aos meus queridos pais, Adauto e Maria Onélia, expresso meu sincero agradecimento por todo o apoio e incentivo, mesmo estando distante fisicamente. Sua confiança em mim foi um estímulo constante para alcançar meus objetivos.

Ao IFTO e à UFPel, agradeço pela oportunidade ímpar concedida, que me permitiu realizar este trabalho e expandir meus horizontes acadêmicos.

Aos professores do PPGC - Doutorado em Computação da UFPel, agradeço pelas valiosas lições e ensinamentos que transcenderam o campo acadêmico.

Aos meus colegas da UFPel e UFRGS-PRODIA, agradeço pelos momentos de descontração, pelo aprendizado, pelas contribuições significativas para este trabalho ao longo desses anos.

À banca examinadora de qualificação e defesa, agradeço por aceitarem o convite e por sua importante contribuição para o desenvolvimento desta tese.

Aos amigos e familiares que estiveram presentes em minha caminhada, expresso minha gratidão. Vocês foram essenciais em cada etapa deste processo.

A todos que, de alguma forma, fizeram parte desta jornada, meus mais sinceros agradecimentos!

Se vi mais longe foi por estar sobre ombros de gigantes.
— SIR ISAAC NEWTON, 1676

RESUMO

MELO, Ramásio Ferreira de. **Um estudo sobre aprendizado de máquina multimodal para identificar a depressão em adolescentes brasileiros usando dados coletados via *smartphones***. Orientador: Ricardo Matsumura Araújo. 2023. 179 f. Tese (Doutorado em Ciência da Computação) – Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, 2023.

A depressão afeta cerca de 322 milhões de pessoas em todo o mundo, segundo a OMS. Os primeiros sintomas da depressão costumam surgir entre o final da adolescência e o início da fase adulta. No entanto, sua apresentação clínica, altamente heterogênea, dificulta o diagnóstico nas fases iniciais, especialmente em ambientes não especializados. Além disso, não existem exames laboratoriais para diagnosticar a depressão. As entrevistas clínicas que seguem critérios operacionais padronizados, como o DSM, são o padrão ouro para o diagnóstico da depressão, apesar do viés subjetivo da avaliação. Neste contexto, as tecnologias móveis, como *smartphones* e sensores vestíveis, oportunizaram novas formas de coletar dados sobre o comportamento depressivo em ambientes domésticos, não observáveis clinicamente, que vão além das abordagens tradicionais. Neste estudo, investigamos modelos multimodais e multitarefas de aprendizado de máquina para prever o humor deprimido atual e futuro dos adolescentes da base de dados do estudo IDEA-RiSCo (*Identifying Depression Early in Adolescence Risk Stratified Cohort*). Realizamos extensos experimentos usando 81 atributos extraídos de dados acústicos, da fala dos adolescentes e do ambiente, acelerômetro e GPS coletados via *smartphones*. Também comparamos o desempenho de modelos preditivos nomotéticos (generalizado) e idiográficos (personalidado) e duas agregações de dados. Os experimentos foram realizados considerando as pontuações do questionário sMFQ como verdade básica. Estabelecemos um modelo de aprendizado de máquina nomotético que combina dados de atividade, áudio de relatos e mobilidade usando agregação cumulativa como uma importante alternativa para complementar o diagnóstico de depressão em adolescentes brasileiros. Além disso, destacamos os atributos que mais contribuíram para otimizar o desempenho dos modelos preditivos, relacionado à atividade: tempo médio diário de caminhada; áudios de relatos: coeficientes MFCC, espectrogramas e contrastes espectrais; e mobilidade: distância máxima de casa e variação de localização.

Palavras-chave: Depressão. Saúde mental. Inteligência artificial. Sensoriamento passivo. Smartphone. Marcadores digitais. Fenotipagem digital.

ABSTRACT

MELO, Ramásio Ferreira de. **A multimodal machine learning study to identify depression in Brazilian adolescents using data collected via smartphones**. Advisor: Ricardo Matsumura Araújo. 2023. 179 f. Thesis (Doctorate in Computer Science) – Technology Development Center, Federal University of Pelotas, Pelotas, 2023.

Depression affects an estimated 322 million people worldwide, according to the WHO. The first symptoms of depression usually appear between late adolescence and early adulthood. However, its highly heterogeneous clinical presentation makes diagnosis difficult in the early stages, especially in non-specialized settings. In addition, there are no laboratory tests to diagnose depression. Clinical interviews and self-reported questionnaires are the gold standard for diagnosing depression, despite the subjective bias of the assessment. In this context, mobile technologies, such as smartphones and wearable sensors, have provided new ways of collecting data on depressed behavior in domestic environments, which are not clinically observable, in addition to traditional approaches. In this study, we investigated multimodal and multitasking machine learning models to predict the current and future depressed mood state of adolescents from the Identifying Depression Early in Adolescence Risk Stratified Cohort (IDEA-RiSCo) database. We performed extensive experiments using 81 features extracted from acoustic data from the adolescents' speech and the environment, accelerometer, and GPS collected through smartphones. Also, we compared the performance of nomothetic (generalized) and idiographic (personalized) predictive models and two data aggregations. The experiments considered the sMFQ questionnaire scores as the ground truth. We established a machine learning model that combines activity data, reported audios, and mobility using cumulative data aggregation as an important alternative to complement the diagnosis of depression in Brazilian adolescents. In addition, we highlight the attributes that most contributed to optimizing the performance of predictive models related to activity: average daily walking time; reported audios: MFCC coefficients, spectrograms, and spectral contrasts; and mobility: maximum distance from home, and location variance.

Keywords: Depression. Mental health. Artificial intelligence. Passive sensing. Smartphone. Digital makers. Digital phenotyping.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

1DCNN	One dimentional Convolutional Neural Network
ARI	Affective Re-activity Index
AUC	Area under Curve
AVEC	Audio/Visual Emotion Challenge and Workshop
BDI	Becks Depression Inventory
BiLSTM	Bidirectional Long Short Time memory
CB	Catboosting
C-SSRS	Columbia-Suicide Severity Rating Scale
CDI	Childrens Depression Inventory
CDRS-R	Childrens Depression Rating Scale Revised
CDRS	Children Depression Rating Scale
CDRS	Childrens Depression Rating Scale
CES-D	Center for Epidemiological Studies Depression Scale
CID	Classificação Internacional de Doenças
CLPSYCH	Computational Linguistics and Clinical Psychology
CNN	Convolutional Neural Network
CTQ	Childhood Trauma Questionnaire
DAIC-WoZ	Distress Analysis Interview Corpus - Wizard of Oz
DAIC	Distress Analysis Interview Corpus
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCT	Discrete Cosine Transform
DSM	Diagnostic and Statistical Manual of Mental Disorders
DNN	Deep Neural Network
E-DAIC	Extended Distress Analysis Interview Corpus
EBM	Eletronic Behavior Monitoring App
EPDS	Edinburgh Postnatal Depression Scale

ET	Extra Trees
F0	Frequência Fundamental
FFT	Fast Fourier transform
FN	False Negative
FP	False Positive
GB	Gradient Boosting
GPS	Global Positioning System
HAMD	Hamilton Rating Scale
HNR	Harmonic to Noise Ratio
IDEA	Identifying Depression Early in Adolescence
IDEA-RS	IDEA Risk Score
IDEA-RiSCo	Identifying Depression Early in Adolescence Risk Stratified Cohort
LLD	Low Level Descriptors
LSTM	Long Short Time memory
MADRS	Montgomery-Asberg Depression Scale
MFCC	Mel-Frequency Cepstral Coefficients
MFQ	Mood and Feelings Questionnaire
MFQ-C	MFQ Adolescent Report
MFQ-P	MFQ Parental Report
MTurk	Amazon Mechanical Turk
MT1DCNN	Multi Task Convolutional Neural Network one dimensional
PHQ	Patient Health Questionnaire
PPV	Positive Predictive Value
RFE	Recursive feature elimination
RSDD	Reddit Self-Reporting Depression Diagnostics
SDS	Zung Self-Rating Depression Scale
SHAPS	Snaith-Hamilton Pleasure Scale
sMFQ	Short Mood and Feelings Questionnaire
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
STFT	Short-time Fourier transform
TNR	True Negative Rate
TN	True Negative

SUMÁRIO

1 INTRODUÇÃO

O Transtorno Depressivo Maior, também conhecido como depressão, é um dos transtornos mentais mais comuns na sociedade. A Organização Mundial da Saúde (OMS) estima que 322 milhões de pessoas atendem aos critérios diagnósticos para depressão (DEPRESSION, 2017). O Brasil, é segundo país com número de casos no continente americano, correspondendo a 5,8% da população e está acima da média global de 4,4%.

A Pesquisa Nacional de Saúde do Brasil (PNS) realizada em 2019 aponta que cerca de 10,2% dos adultos maiores de 18 anos (16,3 milhões de pessoas) sofrem de depressão no país, um aumento de 2,6% desde a última PNS realizada em 2013 (RENDIMENTO, 2014, 2020; STOPA et al., 2020). Além disso, houve aumento global de 18,4% da prevalência da depressão entre os anos de 2005 a 2015 (DEPRESSION, 2017).

O pico de incidência da depressão ocorre entre o fim da adolescência e o início da vida adulta (KIELING et al., 2021). Esse fator contribui para que a depressão esteja entre as maiores causas de morte entre adolescentes no mundo (DICK; FERGUSON, 2015; MULLICK et al., 2022). Além disso, cerca de 10% dos homens e 20% das mulheres sofrem pelo curso crônico da depressão ao longo da vida (LAM, 2018; SHAH et al., 2021).

Os sintomas associados à depressão incluem humor deprimido; anedonia; distúrbios do sono e apetite; retardo psicomotor; fadiga; sentimentos de inutilidade ou culpa excessivos; baixa capacidade de concentração e pensamentos suicidas (APA, 2013).

Atualmente, nenhum exame laboratorial específico orienta o diagnóstico de depressão. As entrevistas clínicas conduzidas por profissionais de saúde que seguem critérios operacionais padronizados são consideradas o padrão ouro para o diagnóstico de depressão, apesar do viés subjetivo inerente à avaliação (LAM, 2018; ASARE et al., 2021).

Estratégias de prevenção e intervenção precoce melhoram os resultados do tratamento de pacientes com depressão (DAVEY; MCGORRY, 2019; THAPAR et al., 2022), porém o monitoramento dos sintomas nas fases iniciais da doença ainda é bastante

difícil, sobretudo em adolescentes (MULLICK et al., 2022).

Além disso, a depressão é uma doença de diagnóstico complexo. Sua apresentação clínica altamente heterogênea causa grande variabilidade nas respostas dos pacientes aos mesmos tratamentos (LAM, 2018; SHAH et al., 2021). Por esse motivo, sobredetecções e subdetecções são desafios comuns na prática clínica (MITCHELL; VAZE; RAO, 2009), especialmente em ambientes não especializados que representam barreiras adicionais ao atendimento adequado.

Estes fatores dificultam o rastreamento das pessoas com sintomas de depressão nas fases iniciais, especialmente em ambientes não psiquiátricos, que associado a demora na busca por tratamento adequado pode torná-lo mais complexo (YUSOF; LIN; GUERIN, 2017).

As tecnologias de hardware e software de *smartphones* e sensores vestíveis tornaram-se capazes de monitorar a saúde, condições ambientais e emocionais das pessoas a um custo bastante reduzido (MASUD et al., 2020). Alguns métodos de coleta de dados mediados por dispositivos móveis foram usados com sucesso no contexto da depressão devido à sua capacidade de agregar dados comportamentais e fisiológicos de pacientes, coletados em seu cotidiano, sobre a evolução dos sintomas e respostas aos tratamentos (SAEB et al., 2015; MASUD et al., 2020; PEDRELLI et al., 2020; MASUD et al., 2020; MOSHE et al., 2021; RYKOV et al., 2021; SHAH et al., 2021; MULLICK et al., 2022).

No entanto, identificar informações comportamentais, não observáveis, mas clinicamente relevantes e incorporá-las aos métodos tradicionais de avaliação baseados em entrevistas psiquiátricas ainda implicam grandes desafios na prática clínica dada a complexidade e heterogeneidade do diagnóstico da depressão.

Além disso, o monitoramento do estilo de vida de pacientes em ambiente doméstico, mediado por tecnologia, é amplamente utilizado para entender o comportamento depressivo em adultos. Contudo, os estudos sobre depressão em crianças e adolescentes são relativamente escassos.

Sistemas baseados em Aprendizado de Máquina (AM) têm evoluído rapidamente e cada vez mais aplicados ao contexto da depressão. Eles conseguiram manipular diferentes categorias de dados, utilizando sensores de *smartphones* e dispositivos vestíveis para investigar a depressão em diferentes grupos demográficos, como, por exemplo, adultos (MASUD et al., 2020; PEDRELLI et al., 2020; ASARE et al., 2021; CHOUDHARY et al., 2022; NGUYEN et al., 2021; YAN; TU; WEN, 2022), gestantes (ZHONG et al., 2022), estudantes universitários (NGUYEN et al., 2021; CHIKERSAL et al., 2021; WARE et al., 2022), usuários de comunidades online (ASARE et al., 2021), pacientes psiquiátricos (HONG et al., 2022), idosos (TAZAWA et al., 2020) e adolescentes (MULLICK et al., 2022) de várias partes do mundo.

Estes sistemas suportam configurações multi-atributos, sendo avaliados em dife-

rentes subtarefas da detecção de depressão com alta capacidade preditiva, para identificar indivíduos deprimidos e não deprimidos (TAZAWA et al., 2020; ASARE et al., 2021; HONG et al., 2022), prever os níveis de depressão (MASUD et al., 2020), prever a intensidade dos sintomas da depressão (PEDRELLI et al., 2020), a identificação precoce de depressão (ZHONG et al., 2022) e em abordagens multi-tarefas de detecção de depressão (CHIKERSAL et al., 2021; MULLICK et al., 2022; CHOUDHARY et al., 2022).

Esta tese avalia a capacidade preditiva de modelos baseados em AM que usam dados coletados via *smartphones* para prever a depressão em adolescentes. Propomos um modelo de AM multimodal capaz de manipular áudios de relatos, áudios episódicos, atividade e mobilidade coletados via *smartphones*.

Nós adotamos o IDEA-RiSCo, banco de dados formado por 150 adolescentes brasileiros com baixo risco, alto risco ou diagnosticados com depressão. Os adolescentes do IDEA-RiSCo interagiram com o chatbot via WhatsApp, responderam a questionários psiquiátricos e tiveram dados dos sensores dos *smartphones* coletados durante duas semanas.

Neste estudo, comparamos diferentes algoritmos de AM em tarefas de classificação binária e regressão, realizamos experimentos extensivos combinando as modalidades de dados, tipos de agregação, e investigamos modelos nomotéticos e idiográficos para prever o humor deprimido dos adolescentes do IDEA-RiSCo.

Além disso, nós adotamos um algoritmo de *Shapley* para destacar os atributos que otimizam o desempenho do modelo desenvolvido e discutimos quais características desses dados podem contribuir mais fortemente como marcadores digitais da depressão.

O avanço tecnológico recente oportunizou novas formas de coletar dados sobre o comportamento depressivo e justificam a importância desta pesquisa dada a complexidade e heterogeneidade da avaliação tradicional da depressão.

Nesse contexto, estudo sobre a depressão realizado em uma amostra nacional, representa uma oportunidade valiosa para compreender o comportamento depressivo na adolescência, e podem auxiliar na elaboração de políticas públicas de saúde mental e programas de prevenção, apoio e intervenção precoce, sobretudo nas fases iniciais da doença.

1.1 Hipóteses

Nesta tese estabelecemos as seguintes hipóteses:

1. Dados coletados via *smartphones* em ambiente doméstico podem discriminar de maneira eficaz os adolescentes com depressão de adolescentes sem depressão.

2. A combinação de atributos acústicos, atividade e/ou mobilidade aumenta o desempenho dos modelos preditivos em comparação com o uso desses atributos individualmente.

1.2 Objetivos de Pesquisa

1.2.1 Objetivo Geral

O objetivo geral desta tese é prever a depressão em adolescentes brasileiros através de dados multimodais coletados via *smartphones*, utilizando algoritmos de aprendizado de máquina.

1.2.2 Objetivos Específicos

Os objetivos específicos desta tese são apresentados a seguir:

1. Avaliar os modelos de AM unimodais e multimodais que consideram atributos acústicos, atividade e mobilidade para prever o humor deprimido atual em uma coorte nacional de adolescentes do IDEA-RiSCo.

2. Comparar o desempenho de modelos nomotéticos e idiográficos, nas atividades de previsão do humor deprimido futuro.

3. Investigar atributos extraídos de áudios de relatos e episódicos, atividade e mobilidade como marcadores digitais de depressão na adolescência.

4. Identificar quais atributos, utilizados individualmente ou em conjunto, têm correlação mais forte com os sintomas da depressão.

1.3 Organização do Trabalho

A tese está estruturada da seguinte forma:

- **Capítulo 2 - Revisão de Literatura**: Este Capítulo apresenta fundamentação teórica sobre a sintomatologia da depressão, descreve alguns instrumentos de avaliação, apresenta o estudo IDEA-RiSCo e técnicas de aprendizado de máquina utilizados nesta tese.

- **Capítulo 3 - Trabalhos relacionados**: Este Capítulo traz uma visão geral sobre os modelos multimodais de AM supervisionados que utilizam dados coletados via *smartphones* aplicados ao contexto da depressão, descreve as fontes de dados, as tarefas de detecção de depressão, marcadores digitais, arquiteturas disponíveis e métricas de avaliação.

- **Capítulo 4 - Metodologia**: Este Capítulo descreve os aspectos metodológicos da pesquisa, a organização dos conjuntos de dados, as estratégias de extração de atributos, métricas utilizadas para avaliar o modelo proposto e os experimentos realizados.
- **Capítulo 5 - Resultados experimentais e Discussão**: Este Capítulo apresenta os resultados dos experimentos desenvolvidos, destaca os atributos que otimizaram o desempenho dos modelos preditivos e analisa a sua correlação com os sintomas da depressão.
- **Capítulo 6 - Conclusão**: Este Capítulo apresenta a conclusão do estudo proposto, as limitações do estudo e os trabalhos futuros.
- **Apêndices**: Por fim, nos apêndices A, B, C, D, E e F são apresentados dados e informações complementares sobre a tese.

2 REVISÃO DE LITERATURA

Este Capítulo apresenta a fundamentação teórica sobre a sintomatologia da depressão, descreve alguns instrumentos de avaliação, apresenta o estudo IDEA-RiSCo e introduz os conceitos e técnicas de aprendizado de máquina utilizados no desenvolvimento desta tese.

2.1 Diagnóstico da depressão

A depressão é um transtorno mental associado ao estigma em múltiplos contextos. Ela pode se manifestar por meio de uma série de sintomas físicos (sono, energia, apetite, libido), emocionais (mau-humor, ansiedade, choro) ou cognitivos (culpa; pessimismo; pensamentos suicidas; problemas de concentração, memória e tomada de decisão) e sua apresentação clínica é bastante variada (KANTER et al., 2008; LAM, 2018).

A incidência da depressão nos jovens aumentou significativamente na última década (THAPAR et al., 2022). Os primeiros sintomas da depressão surgem com maior frequência entre o fim da adolescência e início da vida adulta (KIELING et al., 2019) e estão associados ao aumento da morbidade e da mortalidade em adolescentes e adultos (YUSOF; LIN; GUERIN, 2017; MULLICK et al., 2022). Adicionalmente, a prevalência de depressão ao longo da vida é duas vezes maior em mulheres do que em homens (MALHI; MANN, 2018).

A maioria dos pacientes sofre de fases agudas da doença (episódios) que variam em duração e frequência, com uma probabilidade de recorrência ao longo da vida (MALHI; MANN, 2018). O tratamento da depressão envolve a intervenção medicamentosa e/ou psicoterápica para a remissão dos sintomas na fase aguda e o acompanhamento contínuo desses pacientes para avaliar a possibilidade de recorrência ao longo do tempo (LAM, 2018; KARYOTAKI et al., 2018).

A depressão pode ser categorizada em vários níveis conforme a intensidade dos sintomas, sendo geralmente identificados através do relato subjetivo do paciente ou da observação de outras pessoas. A Figura 1 apresenta, brevemente, as características

que devem ser observadas na sintomatologia da depressão, segundo Apa (2013).

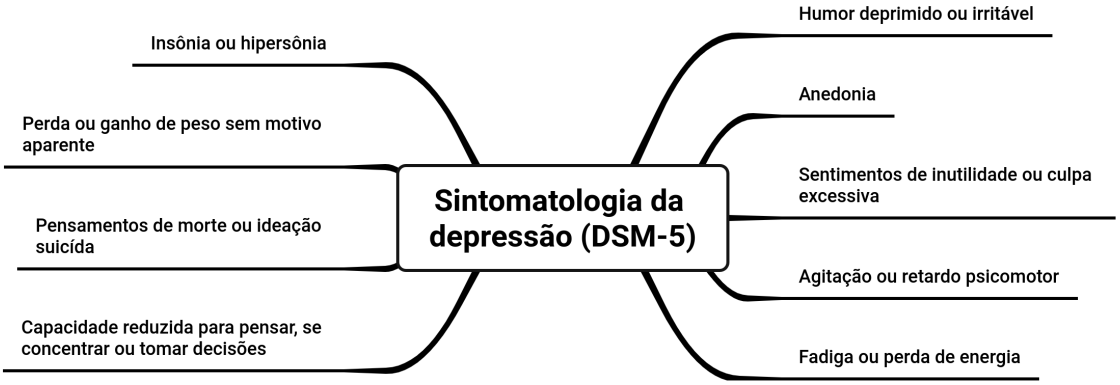


Figura 1 – Sintomatologia da depressão. Fonte: Apa (2013)

Na avaliação clínica, o profissional de saúde considera a frequência em que eles aparecem ao longo do tempo (em dias) em ciclos de pelo menos duas semanas. Um indivíduo tem diagnóstico positivo de depressão quando apresenta pelo menos cinco sintomas; entre eles, deve haver a presença de humor deprimido ou anedonia (APA, 2013; LAM, 2018). Eles são descritos a seguir:

- **Humor deprimido ou irritável**: O humor é frequentemente afetado pela depressão. Pessoas deprimidas manifestam sentimentos intensos de tristeza com mais frequência e por mais tempo do que estados momentâneos de tristeza influenciados por uma determinada situação ou fator externo. As alterações de humor podem se manifestar de maneira diferente em crianças e adolescentes, apresentando comportamentos mais retraídos, ou sentimentos de raiva intensa e irritabilidade.

- **Anedonia**: Uma acentuada diminuição do interesse ou prazer manifesta-se na forma de tédio ou sentimentos de indiferença e comumente pode prejudicar o desempenho das atividades diárias, interações sociais e relacionamentos conjugais devido à perda ou diminuição da libido;

- **Sentimentos de inutilidade ou culpa excessiva**: São comuns em pacientes deprimidos que podem reagir negativamente a estímulos favoráveis ou eventos triviais do cotidiano.

- **Agitação ou retardo psicomotor**: Pessoas com depressão grave geralmente apresentam retardo psicomotor que pode causar lentidão dos movimentos corporais, falta de expressão facial, fala reduzida e espaçada, entre outros prejuízos às habilidades físicas e mentais.

- **Capacidade reduzida para pensar ou concentrar-se ou tomar decisões**: A dificuldade de pacientes deprimidos em manter a concentração pode levar a problemas cognitivos e dificuldades de memória e comprometer a tomada de decisões em tarefas importantes.

- **Pensamentos de morte ou ideação suicida**: Sentimentos de desesperança podem dar origem a pensamentos de morte, muitas vezes vistos como uma forma de aliviar o sofrimento. Ideação ou comportamentos suicidas ocorrem em quase dois terços dos pacientes com depressão grave.

- **Fadiga ou perda de energia**: A sensação de mal-estar constante, com desgaste físico e mental, também é comum e pode prejudicar o desenvolvimento pessoal e profissional de pessoas com depressão. Em casos graves, comprometem as atividades diárias básicas, como a higiene pessoal.

- **Perda ou ganho de peso sem motivo aparente**: A depressão pode causar distúrbios alimentares e alterações físicas nas pessoas. Tanto o ganho quanto a perda de peso súbito, sem motivo aparente ou em excesso (com menos frequência) podem ser observados na sintomatologia da depressão;

- **Insônia ou hipersonia**: Alterações do sono também são comuns em pacientes com depressão. A maioria dos pacientes deprimidos têm dificuldade em manter um sono regular e estável. É comum acordar várias vezes durante a noite ou ter dificuldade para dormir ao ir para a cama quando a depressão está associada à ansiedade. Embora casos de hipersonia também possam ser característicos de depressão.

O manejo clínico da depressão envolve as etapas de triagem, avaliação, desenvolvimento de uma aliança terapêutica, seleção do(s) tratamento(s), monitoramento e acompanhamento (LAM, 2018).

Nenhum exame laboratorial específico orienta o diagnóstico de depressão. Em vez disso, a avaliação da depressão considera a história clínica e o exame do estado mental do paciente. Ela é baseada na experiência de psiquiatras ou outros profissionais de saúde, entrevistas, questionários e outras técnicas de avaliação comportamental baseadas em critérios operacionais padronizados, como o DSM. (MALHI; MANN, 2018; CHEVANCE et al., 2020).

As entrevistas clínicas conduzidas por profissionais de saúde e a aplicação de questionários autorrelatados são reconhecidamente os métodos mais confiáveis para o diagnóstico de depressão usados atualmente.

2.1.1 Instrumentos classificatórios da depressão

Os principais instrumentos classificatórios de depressão são o Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM - do inglês *Diagnostic and Statistical Manual of Mental Disorders*) (APA, 2013) e a Classificação Internacional de Doenças (CID-10), com predominância do DSM na pesquisa científica (MALHI; MANN, 2018).

Os questionários são as principais ferramentas de captação de informações sobre os sintomas da depressão. Eles são utilizados há décadas, em vários estágios do manejo clínico, para avaliar a intensidade dos sintomas e remissão, na atenção primária ou em ambientes especializados (MALHI; MANN, 2018). Além disso, questionários psiquiátricos e testes adaptativos computadorizados são amplamente utilizados para triagem populacional de depressão em ambientes não psiquiátricos (GRAHAM et al., 2019; GIBBONS; DEGRUY, 2019).

Os questionários podem ser autorrelatados ou conduzidos por profissionais especializados. Eles podem ser direcionados a diferentes perfis sociodemográficos, também variam no número de questões, indicadores e níveis de intensidade dos sintomas. A gravidade da depressão é avaliada de forma diferente conforme o questionário adotado, mas, em geral, escores mais altos significam sintomas mais graves e maior probabilidade de o indivíduo estar deprimido.

Alguns questionários foram desenvolvidos para avaliar a depressão em um contexto geral como, por exemplo:

- Escala de Depressão do Centro de Estudos Epidemiológicos (CES-D - do inglês *Center for Epidemiological Studies Depression Scale*) desenvolvido por Radloff (1977);
- Inventário de Depressão de Beck (BDI - do inglês *Beck's Depression Inventory*) criado em Beck; Steer; Brown (1996);
- Escala Zung de Auto Avaliação de Depressão (SDS - do inglês *Zung Self-Rating Depression Scale*) desenvolvido por Zung; Richards; Short (1965);
- Questionário de Saúde do Paciente (PHQ - do inglês *Patient Health Questionnaire*) concebido por Kroenke et al. (2009);
- Escala de Avaliação de Hamilton (HDRS, HAMD - do inglês *Hamilton Rating Scale*) desenvolvido em Hamilton (1960);
- Escala de Depressão de Montgomery-Asberg (MADRS - do inglês *Montgomery-Asberg Depression Scale*) proposto por Montgomery; Åsberg (1979);

Outros podem ser aplicados em contextos específicos como a Escala de Depressão Pós-parto de Edimburgo (EPDS - do inglês *Edinburgh Postnatal Depression Scale*)

criado por Cox; Holden; Sagovsky (1987) para rastrear sintomas de depressão em mulheres até a oitava semana após o final da gravidez; e a Escala de classificação de depressão infantil revisada (CDRS-R - do inglês *Children's Depression Rating Scale Revised* proposta em Poznanski; Mokros (1996) para avaliar a depressão infantil, em crianças de 6 a 12 anos.

A Tabela 1 apresenta alguns dos principais questionários utilizados para investigar sintomas de depressão em adultos.

Tabela 1 – Instrumentos de avaliação dos sintomas da depressão em indivíduos adultos.

Escala	Condução	Item	Pontuação Geral				
			Normal	Leve	Moderada	Mod. Grave	Grave
CES-D	Autorrelato	20	0-16	-	16-23	-	24-60
EPDS	Autorrelato	10	0-8	9-11	12-13	-	14-30
BDI	Autorrelato	21	0-13	14-19	20-28	-	29-63
SDS	Autorrelato	20	20-44	45-59	60-69	-	70-80
PHQ	Autorrelato	8/9	0-4	5-9	10–14	15-19	20-
HAMD	Clínico	17/21	0-7	8-13	14-18	19-22	23-
MADRS	Clínico	10	0-6	7-19	20-34	-	35-54

Além disso, existem alguns questionários criados especificamente para auxiliar o diagnóstico da depressão em crianças e adolescentes. Esses questionários são brevemente descritos em seguida:

- *Patient Health Questionnaire Depression Scale* (PHQ): questionário de 8 ou 9 itens (PHQ-8/9), bastante usado para rastrear a depressão na atenção primária. Possui cinco níveis de intensidade dos sintomas. O diagnóstico de depressão é considerado positivo a partir do segundo ciclo consecutivo se o paciente obtiver entre 5 e 9 pontos para depressão leve, 10 a 14 pontos para depressão moderada, 15 a 19 pontos para depressão moderadamente grave e a partir de 20 pontos para depressão grave. A versão PHQ-A é utilizada para avaliar os sintomas da depressão em adolescentes.

- *Beck's Depression Inventory* (BDI): questionário autoavaliativo publicado inicialmente em 1961 e revisado em 1996 (BDI-II) que pode ser aplicado em adolescentes e adultos. O questionário BDI é formado por 21 questões que abordam sentimento de culpa, condições de sono, antipatia, autocrítica e outros. Seus escores variam de 0 a 63, categorizados em normal (0-13), leve (14-19), moderada (12-28) e grave (29-63).

- *Mood and Feelings Questionnaire (MFQ)*: é um questionário desenvolvido para o estudo de sintomas da depressão na infância e adolescência formado por 33 itens. Ele pode ser usado para crianças e adolescentes com idades entre 8

e 17 anos. Existem três versões do MFQ: O *Short-MFQ* que é uma versão reduzida do MFQ de 13 itens; o *MFQ - Adolescent Report* (MFQ-C) usado para o adolescente responder sobre si e o *MFQ - Parental Report* (MFQ-P) no qual os pais ou responsáveis legais avaliam a sintomatologia do adolescente (ROSA et al., 2018).

- *Children Depression Rating Scale (CDRS)*: foi desenvolvida por Poznanski; Cook; Carroll (1979) e revisada (CDRS-R) em Poznanski; Mokros (1996), para avaliar a gravidade e remissão dos sintomas da depressão em crianças de 6 a 12 anos. É um questionário formado por 17 itens, conduzido por um clínico com intervalo de pontuação entre 17 e 113. Uma pontuação $\geq$ 40 é indicativa de depressão, enquanto uma pontuação $\leq$ 28 indica ausência de sintomas ou remissão. O CDRS-R foi originalmente concebido para avaliar a depressão infantil, mas também é comumente usado em adolescentes (MAYES et al., 2010).

- *Children's Depression Inventory (CDI)*: é um questionário derivado do BDI desenvolvido por Kovacs (1992). Ele é usado como ferramenta de triagem da depressão em crianças e adolescentes com idades entre 8 e 17 anos. O questionário CDI contempla 5 domínios da depressão: Anedonia, Ineficácia, Problemas Interpessoais, Humor Negativo e Autoestima Negativa. Ele é composto por 27 itens com pontuações de 0 a 2 para cada item. Sua pontuação pode variar de 0 a 54, sendo 20 o escore de corte indicado para rastreamento de depressão.

Existem ainda questionários que podem ser usados para complementar o diagnóstico da depressão. Como por exemplos:

- Escala de classificação de gravidade do suicídio de Columbia (C-SSRS - do inglês *Columbia-Suicide Severity Rating Scale*) proposto em Posner et al. (2008), conduzido por um profissional de saúde para investigar ideação e comportamento suicida;

- Questionário de Trauma na Infância (CTQ - do inglês *Childhood Trauma Questionnaire*), questionário autorrelatado criado por Bernstein et al. (1998) para adolescentes e adultos que investiga o histórico de abuso e negligência na infância;

- Índice de Reatividade Afetiva (ARI - do inglês *Affective Reactivity Index*) proposto por Stringaris et al. (2012) para medir o grau de irritabilidade do paciente na infância e adolescência;

- Escala de Prazer Snaith-Hamilton (SHAPS - do inglês *Snaith-Hamilton Pleasure Scale*) desenvolvido por Snaith et al. (1995) para avaliar a presença de anedonia.

Os questionários descritos anteriormente são ferramentas tradicionais da psiquiatria usadas para coletar informações sobre os sintomas da depressão. Entretanto, tecnologias de hardware e softwares como os *smartphones*, sensores vestíveis, redes sociais, estão oportunizando novas formas de aquisição de dados de pacientes deprimidos que ultrapassam os limites da observação profissional em ambiente especializado e permitem gerar hipóteses biologicamente testáveis sobre a sintomatologia da depressão.

2.2 Estratégias para coleta de dados de pessoas com depressão

Não existe, até onde sabemos, uma base de dados padrão-ouro para pesquisas que utilizam algoritmos de AM no contexto da depressão. Em geral, a manipulação de dados de pessoas com depressão está sujeita a restrições éticas e legais. Além disso, é fundamental que os instrumentos de avaliação psiquiátrica sejam usados para validar os conjuntos de dados para torná-los mais confiáveis.

Esses são alguns dos fatores que tornam a criação de bases de dados de depressão mais desafiadora. Contudo, identificamos na literatura, algumas estratégias, utilizadas com sucesso para obter dados de pessoas com depressão.

Alguns estudos usaram a força de trabalho disponível em plataformas de *crowdsourcing* como Qualtrics e *Amazon Mechanical Turk* (MTurk) para construir conjuntos de dados com grande número de participantes e sob várias abordagens, por exemplo, para coletar dados de usuários de redes sociais, aplicar questionários de saúde mental e validação de dados por profissionais de saúde (SCHWARTZ et al., 2014; REECE; DANFORTH, 2017; SHUAI et al., 2017; GUNTUKU et al., 2019). No entanto, esses dados são coletados em ambientes heterogêneos, bastante ruidosos e com pouca padronização.

Dados de pessoas com depressão também foram coletados por meio de questionários para detectar depressão pós-parto conforme o estudo desenvolvido em Natarajan et al. (2017), ou por entrevistas com pacientes previamente diagnosticados com depressão em clínicas psiquiátricas em análises conduzidas por Doryab et al. (2014) e Anis et al. (2018).

Alternativamente, alguns eventos científicos fornecem bases de dados para o desenvolvimento de sistemas de detecção de depressão. A saber:

- O *The Audio/Visual Emotion Challenge and Workshop* (AVEC) disponibiliza aos participantes, a base de dados de entrevistas clínicas (*Distress Analysis Interview Corpus* - DAIC) (2014) ou suas variações *Distress Analysis Interview Corpus - Wizard of Oz* (DAIC-WoZ) de 2017, *Extended Distress Analysis Interview Corpus* (E-DAIC) de 2019. Os conjuntos de dados são disponibilizados nos formatos de texto, áudio e vídeo, e validados por questionários BDI-II na versão de

2014 e PHQ-8 nas versões de 2017 e 2019.

- O *Erisk* é uma competição anual que disponibiliza aos participantes a base de dados extraída da rede social Reddit, validada pelo questionário BDI.
- O *Workshop* de Linguística Computacional e Psicologia Clínica (CLPSYCH - do inglês Computational Linguistics and Clinical Psychology) também disponibiliza uma base de dados de usuários do Reddit, o Reddit Self-Reporting Depression Diagnostics (RSDD). No entanto, o RSSD não utiliza nenhum questionário psiquiátrico para validação das amostras.
- O *StudentLife* (WANG et al., 2014) inclui dados de 48 estudantes universitários, 38 homens e 10 mulheres, do *Dartmouth College*. O *StudentLife* contém dados coletados passivamente via *smartphones* Android: atividade física, áudio, conversas, *Bluetooth*, sensor de luz, localização por GPS e Wi-Fi, telefone, bloqueio de telefone e Wi-Fi. Os participantes foram monitorados por dez semanas e completaram o questionário PHQ-9 usado para validar o conjunto de dados.

A Tabela 2 apresenta informações sobre os principais conjuntos de dados disponíveis para pesquisas de detecção de depressão encontradas na literatura (MUNDT et al., 2007; VALSTAR et al., 2013; ALGHOWINEM, 2013; GRATCH et al., 2014; WANG et al., 2014; LOSADA; CRESTANI, 2016; YATES; COHAN; GOHARIAN, 2017; YANG et al., 2017; WILLIAMSON et al., 2019; CAI et al., 2020).

Essas bases de dados foram coletadas organizadamente, em ambiente controlado e constituem um relevante ecossistema de suporte para pesquisas e desenvolvimento de soluções tecnológicas baseadas em AM aplicadas à depressão. O acesso a esses bancos de dados é permitido mediante autorização prévia de seus mantenedores. Os pesquisadores podem usar esses conjuntos de dados para fins de pesquisa científica e não comercial. Todos os conjuntos de dados contêm dados de indivíduos adultos.

Tabela 2 – Bases de dados de depressão

Bases	Idioma	Origem	Modalidade	Participante (H./M.)	Versões	Validação	Referência
Mundt	Inglês	Entrevista	Áudio, Vídeo	35 (15/20)	-	HAM-D	Mundt et al. (2007)
AViD	Alemão	Entrevista	Áudio, Vídeo	32 (9/23)	-	BDI-II	Valstar et al. (2013)
BlackDog	Inglês (AU)	Entrevista	Áudio, Vídeo	60 (30/30)	-	QIDS-SR	Alghowinem (2013)
Pitt	Inglês	Entrevista	Áudio, Vídeo	38 (14/24)	-	HRSD	Yang; Fairbairn; Cohn (2012)
DAIC	Inglês	Entrevista	Texto, Áudio e Vídeo	621	DAIC-WOZ, E-DAIC	BDI-II	Gratch et al. (2014)
Erisk	Inglês	*Reddit*	Texto	90	-	BDI-II	Losada; Crestani (2016)
RSDD	Inglês	*Reddit*	Texto	116,484	RSDD-Time	-	Yates; Cohan; Goharian (2017)
WIBD	Inglês	Entrevista	Áudio, Vídeo	18	-	HAM-D	Williamson et al. (2019)
Modma	Chinês	Entrevista	Áudio, EEG	52 (36/16)	-	PHQ-9	Cai et al. (2020)
StudentLife	Inglês	*Smartphone*	Luz, *Bluetooth*, Áudio, Wi-Fi, tela, uso de aplicativos, sono, atividade, GPS	48 (38/10)	-	PHQ-9	Wang et al. (2014)

2.3 *Identifying Depression Early in Adolescence (IDEA)*

O *Identifying Depression Early in Adolescence* (IDEA) é um consórcio interdisciplinar global que reúne pesquisadores de diversas áreas do conhecimento de países como África do Sul, Brasil, Estados Unidos, Nepal, Nigéria e o Reino Unido, visando conduzir uma extensa investigação sobre características neurobiológicas associadas ao risco de desenvolver depressão e à presença de depressão na adolescência (KIELING et al., 2019).

No Brasil foram avaliados 7.720 adolescentes com idade entre 14 e 16 anos, estudantes de 101 escolas públicas estaduais da cidade de Porto Alegre, no Rio Grande do Sul. Na etapa inicial de triagem escolar, os adolescentes responderam a um questionário PHQ-A e a um questionário *IDEA Risk Score* (IDEA-RS) proposto no estudo de Rocha et al. (2021), para estimar o risco de depressão precoce na adolescência.

O IDEA-RS é um modelo de prognóstico multivariável aplicado em um estudo longitudinal de 3 anos para avaliar o risco de adolescentes com idade inicial de 15 anos desenvolverem a depressão aos 18 anos. Ele foi desenvolvido a partir de dados da Coorte de Nascimentos, ano 1993, da cidade de Pelotas, no Rio Grande do Sul e foi validado em coortes do Reino Unido e Nova Zelândia. O IDEA-RS utiliza informações sociodemográficas como sexo; cor da pele; uso de drogas; fracasso escolar, isolamento social; envolvimento em brigas; relacionamento com os seus pais e entre eles; maus-tratos na infância e fuga de casa para estimar o risco de depressão em adolescentes (KIELING et al., 2021; ROCHA et al., 2021).

Além disso, foi estabelecida uma nova coorte: a *Identifying Depression Early in Adolescence Risk Stratified Cohort (IDEA-RiSCo)* que vem acompanhando em profundidade 150 dos 7.720 adolescentes que atenderam aos critérios estabelecidos na etapa de triagem escolar. São 50 adolescentes com baixo risco para depressão, 50 adolescentes com alto risco para depressão e 50 adolescentes com diagnóstico de depressão. Os adolescentes considerados com alto e baixo risco para depressão não foram diagnosticados previamente com a doença, enquanto os participantes com depressão contemplam os critérios para o diagnóstico positivo, porém nunca fizeram tratamento (KIELING et al., 2021).

O IDEA-RiSCo é um estudo sobre medidas neurobiológicas, psicológicas e ambientais que incluem avaliação clínica detalhada, entrevistas com os adolescentes e responsáveis legais, coleta de amostras de sangue e saliva, exames de ressonância magnética e neuroimagem, coletas de dados remota via *smartphones*, entre outros recursos tecnológicos visando o desenvolvimento de estratégias de detecção precoce de depressão, bem como sua prevenção na adolescência (KIELING et al., 2019, 2021).

O estudo IDEA-RiSCo é organizado em quatro períodos de coleta de dados, chamados de ondas (W - do inglês Waves). A Tabela 3 apresenta as etapas desenvolvidas

no estudo IDEA-RiSCo em cada uma das ondas.

Tabela 3 – Etapas e ondas do estudo IDEA-RiSCo

Ondas	W0	W1	W2	W3
Triagem	x			
Questionários online	x	x	x	x
Avaliação clínica	x			x
Coleta de material biológico	x			x
Coleta de dados remota via *smartphones*	x	x	x	x
Coleta de marcadores cronobiológicos	x			x

A onda 0 (linha de base) ocorreu no ano de 2018, envolve as etapas de triagem dos adolescentes, avaliação clínica no Centro de Pesquisa Clínica do Hospital de Clínicas de Porto Alegre (HCPA) e coleta de dados remota realizada através dos *smartphones* dos adolescentes. As ondas 1 e 2 ocorreram remotamente, um e dois anos após a avaliação de linha de base, respectivamente. Os adolescentes e responsáveis legais responderam questionários online da psiquiatria e tiveram os dados de seus *smartphones* coletados remotamente. A onda 3, ocorreu no terceiro ano após a avaliação da linha de base. Envolve a aplicação de questionários da psiquiatria para adolescentes e responsáveis legais, avaliação clínica, coleta de material biológico, exames de ressonância magnética, coleta de dados remota via *smartphones* e de marcadores cronobiológicos. Mais detalhes sobre o processo de coleta de dados são apresentados em seguida.

- **Triagem**: Durante a etapa de triagem, 7.720 adolescentes estudantes de escolas públicas de Porto Alegre-RS responderam os questionários PHQ-A e IDEA-RS. Além disso, foram coletadas informações como nome, data de nascimento, sexo, raça/cor de pele, informações de contato dos adolescentes e responsáveis legais, entre outros dados sociodemográficos.
- **Questionários online**: Questionários foram aplicados aos adolescentes e seus responsáveis legais em todas as ondas. Além do PHQ-A e IDEA-RS usados em W0 na etapa de triagem. Nas ondas 1 e 2 os adolescentes responderam aos questionários online:
  - Questionário de Saúde do Paciente - Adolescente (PHQ-A - do inglês *Patient Health Questionnaire for Adolescents*)
  - DSM-5 Auto Avaliação Nível 1 Medida de Sintomas Transversais - Criança (CCSM-C - do inglês *DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure, Child*)
  - Questionário de Humor e Sentimentos - Criança (MFQ-C - do inglês *Mood and Feelings Questionnaire for Child/Adolescents*).

Os responsáveis legais responderam os questionários:

- Questionário de Humor e Sentimentos - Pais (MFQ-P - do inglês *Mood and Feelings Questionnaire for Parents*)
- Medida Transversal de Sintomas de Nível 1 do DSM-5 - Pais (CCSM-P - do inglês *Cross-Cutting Symptom Measure, Parent*)

Na onda 3, os adolescentes responderam os questionários das ondas 1 e 2, acrescido dos questionários a seguir:

- Critério Brasileiro de Classificação Econômica (ABEP)
- Pesquisa de Saúde e Impacto do Coronavírus (CRISIS),
- Teste de Triagem para Envolvimento com Álcool, Tabagismo e Substâncias (ASSIST - do inglês *The Alcohol, Smoking and Substance Involvement Screening Test*)
- Instrumento de Vínculo Parental (PBI - do inglês *Parental Bonding Instrument*),
- Questionário de Trauma na Infância (CTQ - do inglês *Childhood Trauma Questionnaire*),
- Escala de Prazer Snaith-Hamilton (SHAPS - do inglês *Snaith-Hamilton Pleasure Escale*),
- Índice de Reatividade Afetiva (ARI - do inglês *Affective Reactivity Index*),
- Escala de Ansiedade Infantil de Spence Versão Infantil (SCAS-C - do inglês *Spence Children's Anxiety Scale - Child version*),
- Inventário de Força Juvenil - Autorrelato de Adolescente (YSI-A - do inglês *Youth Strength Inventory – Adolescent version*),
- Escala de Resiliência Adaptada (ARS),
- Escala de Características de Personalidade Borderline para Crianças 11 (BPFSC-11 - do inglês *Borderline Personality Feature Scale for Children – 11*).

Os responsáveis legais responderam os questionários das ondas 1 e 2, acrescidos dos seguintes questionários:

- Índice de Reatividade Afetiva - Pais (ARI-P - do inglês *Affective Reactivity Index* - Parents version),
- Inventário de Força Juvenil - Pais (YSI-P - do inglês *Youth Strength Inventory – Parents version*),

  - Escala de Ansiedade Infantil Spence - Pais (SCAS-P - do inglês *Spence Children's Anxiety Scale - Parents version*),
  - Questionário de Humor e Sentimentos - Autorrelato dos Pais (MFQ-A - do inglês *Mood and Feelings Questionnaire - Adult self-report*).

- **Avaliação clínica**: A etapa de avaliação clínica ocorre nas ondas 0 e 3. Ela foi conduzida por psiquiatras de crianças e adolescentes do HCPA. Durante a avaliação, os profissionais entrevistaram os adolescentes e seus responsáveis legais individualmente e aplicaram o questionário Programação para Transtornos Afetivos e Esquizofrenia para Crianças em Idade Escolar-Presente e Versão Vitalícia (K-SADS-PL - do inglês *Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version*).

- **Coleta de material biológico**: Após a avaliação clínica, medidas antropométricas como altura, peso, circunferência da cintura e do quadril e temperatura axilar; amostras de sangue e saliva; e exames de ressonância magnética foram coletados dos adolescentes. Todo o processo de coleta foi realizado por pesquisadores e profissionais de saúde especializados do HCPA.

- **Coleta de dados remota via smartphone**: Dados de diferentes modalidades foram coletados remotamente em todas as ondas através dos aplicativos *Google Location History* (GLH) e *Google Fit App* (GFIT), instalados nos *smartphones* dos adolescentes. Nas ondas 2 e 3 os dados dos adolescentes foram coletados através dos aplicativos *Electronic Behavior Monitoring* (EBM) versão 2.0, Brain explorer e o IDEA-Bot. Esse conjunto de dados dos adolescentes coletados via *smartphones* integram um projeto *spin-off* denominado IDEA-RiSCo, estabelecido com apoio da *Royal Academy of Engineering* do Reino Unido e viabilizou o desenvolvimento desta tese.

- **Coleta de marcadores cronobiológicos**: Nesta etapa, foram utilizados aparelhos que medem a actimetria de pulso para monitorar o sono, a exposição à luz e a movimentação dos adolescentes. Além disso, os adolescentes e seus responsáveis responderam questionários como Índice de Higiene do Sono (SHI - do inglês *Sleep Hygiene Index*), Índice de Qualidade do Sono de Pittsburgh (PSQI - do inglês *Pittsburgh Sleep Quality Index*) entre outros questionários que avaliam a qualidade do sono. Todos esses dados compõem o estudo Chrono-IDEA.

Os requisitos necessários para garantir a privacidade e a confidencialidade foram previamente discutidos com os participantes do projeto e seus responsáveis.

2.3.1 Base de dados IDEA-Tech

O IDEA-Tech utiliza métodos de coleta de dados ativa e passiva. Ao longo de um período de 15 dias, 150 adolescentes tiveram seus dados monitorados diariamente, das 6:00 às 23:59, utilizando aplicativos específicos instalados em seus *smartphones Android*. Os aplicativos usados para a coleta de dados foram IDEA-Bot, EBM, GFIT, GLH e Brain Explorer. A Figura 2, apresenta a base de dados do IDEA-Tech bem como suas diferentes modalidades.

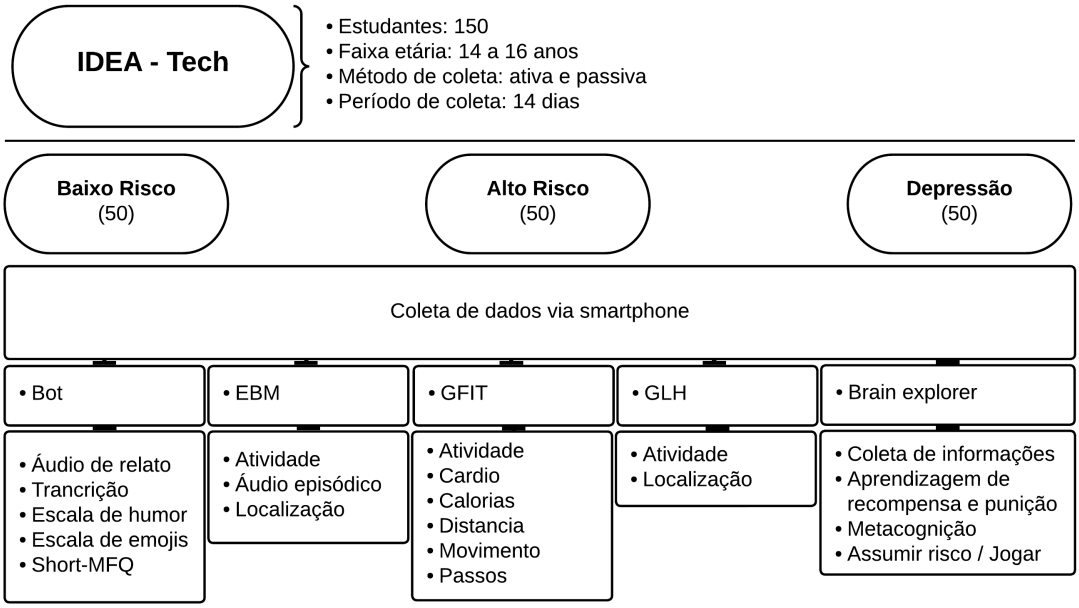


Figura 2 – Base de dados IDEA-Tech

A base de dados IDEA-Tech inclui o código de identificação dos adolescentes, informações sobre sexo e grupos de risco; áudios enviados pelos adolescentes e capturados do ambiente; dados captados através de acelerômetros e GPS; escalas de humor e questionários psiquiátricos preenchidos pelos adolescentes e seus responsáveis; data e hora em que os dados foram coletados. Os aplicativos de coleta e os dados disponíveis no IDEA-Tech são detalhados em seguida.

- **Record_id**: Código de identificação exclusivo e não rastreável do adolescente.
- ***Timestamp***: O momento (dia, mês, ano, hora, minuto e segundo) em que os dados foram coletados pelos aplicativos e questionários respondidos pelos adolescentes.
- **Grupos de Risco**: Os adolescentes que participaram da coleta de dados foram divididos igualmente em três grupos (Baixo Risco, Alto Risco, Depressão) conforme o risco de desenvolverem a depressão durante o estudo IDEA-RiSCo.

- **Sexo**: No IDEA-RiSCo os adolescentes são distribuídos de forma igualitária por sexo entre os grupos. Isso significa que, para os grupos de baixo e alto risco e com depressão, metade dos adolescentes é do sexo feminino e outra metade do sexo masculino.

O IDEA-Bot (VIDUANI et al., 2023) é um *chatbot* desenvolvido especificamente para realizar a coleta ativa de dados no IDEA-RiSCo. Ele capturou informações textuais sobre o humor e amostras voluntárias da fala dos adolescentes, uma vez ao dia e em dias alternados, entre 8:00 e 23:59 durante 15 dias.

Especificamente, os adolescentes foram incentivados a responder perguntas relacionadas ao seu cotidiano por meio de áudios (áudios de relatos) e preencher questionários de humor em dois formatos diferentes: um baseado em escalas de 0 a 10 e outro usando cinco opções de emojis. Toda a interação com os adolescentes, assim como as informações coletadas, foram gerenciadas pelo IDEA-Bot através do *WhatsApp*. Os dados coletados a partir da interação dos adolescentes com o IDEA-Bot são apresentados em seguida.

- **Áudios de relatos**: São amostras voluntárias da fala dos adolescentes, coletadas em tempo real via *WhatsApp*. O IDEA-Bot tinha o objetivo de coletar ao menos 1 minuto de áudio dos adolescentes por dia, sem limites para duração máxima, enviadas em resposta a várias perguntas disponibilizadas durante o período de coleta de dados. Nos dias 1, 3, 5, 7, 9, 11, 13, em quatro momentos entre as 8:00 e 23:59, os adolescentes receberam mensagens de texto solicitando que eles gravassem áudios em respostas. Algumas perguntas enviadas pelo IDEA-Bot são apresentadas a seguir:
  - Dia 1. Pergunta 3: O que você fez hoje? Hoje é um dia normal de rotina?
  - Dia 3. Pergunta 1: O que você está fazendo agora? Você está com alguém?
  - Dia 5. Pergunta 2: E como está seu bairro? Há coisas boas por aí?
  - Dia 7. Pergunta 1: Hoje eu quero saber sobre sua história favorita. Pode ser um filme, uma série, um livro... O que você quiser!
  - Dia 9. Pergunta 1: Você usa muito o telefone? O que você mais gosta de fazer nele?
  - Dia 11. Pergunta 1: Além de mandar áudios aqui, com quem você conversa sobre as coisas que acontecem na sua vida? Como está sua relação com essa pessoa?
  - Dia 13. Pergunta 2: E como foram essas duas últimas semanas? Aconteceu algo diferente?

Finalmente, as amostras de áudio contendo os relatos dos adolescentes são armazenadas em formato OGA.

- **Transcrições**: Os áudios de relatos foram transcritos pela equipe de pesquisadores do IDEIA-RiSCo. Ao menos três pesquisadores estiveram envolvidos no processo de transcrição dos áudios de relatos dos adolescentes. As informações pessoais ou quaisquer outras consideradas sensíveis presentes nos áudios de relatos, dos adolescentes ou de terceiros, foram codificadas ou omitidas para não permitir a identificação dos participantes. Atualmente, as transcrições estão disponíveis somente no idioma português.

- **Questionários de avaliação do humor usando escala likert**: Os adolescentes preencheram, diariamente, uma escala numérica para avaliar o humor que varia entre 0 (muito triste) a 10 (muito feliz).

- **Questionários de avaliação do humor com interação por emojis**: Outra escala para avaliar o humor dos adolescentes que contém cinco opções de emojis.

Em paralelo, foi realizada uma coleta passiva de dados usando os aplicativos EBM, GFIT e GLH, instalados nos *smartphones Android* dos adolescentes. Esses aplicativos são iniciados automaticamente sempre que o dispositivo é ligado, mas em qualquer tempo, os adolescentes puderam interromper a coleta de dados.

O EBM v2.0 (MAHARJAN et al., 2021) coleta os dados dos *smartphones* dos adolescentes com uma frequência de uma vez a cada 15 minutos. Essa coleta inclui coordenadas do *Global Positioning System* (GPS); medidas do acelerômetro interpretadas pelo *Android* e áudios capturados do ambiente onde os adolescentes estão inseridos. Os dados coletados pelo EBM são detalhados em seguida.

- **Atividade**: É avaliada através dos dados do acelerômetro de 3 eixos, contido nos *smartphones* dos adolescentes. Eles são analisados por uma *API Activity Recognition* que detecta automaticamente as atividades que os adolescentes estão realizando no momento da captura. Existem seis atividades possíveis: "*Still, Unknown, On_Foot, Tilting, In Vehicle, In Bicycle*". Cada amostra dos dados de atividade inclui o identificador do participante; a atividade, um valor de confiança e o momento em que o dado foi coletado.

- **Mobilidade**: É avaliada através das coordenadas do GPS captadas usando o *smartphone* dos adolescentes. Cada amostra dos dados de GPS inclui o identificador do participante; coordenadas de altitude, latitude, longitude e um valor de Precisão; a data e hora da coleta.

- **Áudios episódicos**: São áudios com duração entre 5 e 30 segundos coletados dos *smartphones* dos adolescentes para avaliar as interações sociais positivas e negativas no seu ambiente doméstico. Neste estudo, esses áudios episódicos foram transformados em variáveis numéricas associadas à presença e ausência de fala humana.

O GFIT pratica a coleta de dados de atividades físicas realizadas pelos adolescentes em janelas de 15 minutos. Os atributos extraídos usando o GFIT, disponíveis no IDEA-Tech, são calculados automaticamente usando a agregação diária dos dados. São eles: contagem de minutos em movimento; calorias; distância percorrida em metros; pontos de cardio; minutos de cardio; velocidade média; velocidade máxima; velocidade mínima; contagem de passos e duração das atividades: bicicleta, caminhada e corrida em milissegundos.

O GLH realiza a coleta de dados de GPS (latitudes e longitudes), e dados de atividades (a atividade e o momento inicial e final de cada atividade). As atividades disponíveis no GLH coletadas dos *smartphones* dos adolescentes do IDEA-RiSCo são: *"CYCLING, FLYING, IN_BUS, IN_PASSENGER_VEHICLE, IN_SUBWAY, IN_TRAIN, MOTORCYCLING, UNKNOWN, WALKING"*.

A Tabela 4 apresenta o número de adolescentes cujos dados foram coletados passivamente em cada onda durante o estudo IDEA-RiSCo.

Tabela 4 – Coleta de dados passiva do estudo IDEA-RisCo

App	Onda	GPS	Atividade	Episódicos
GFIT	W3	-	74	-
GLH	W0	24	23	-
	W1	32	29	-
	W2	33	50	-
	W3	51	50	-
EBM	W2	105	104	102
	W3	90	84	63

Vale ressaltar que algumas etapas do estudo IDEA-RiSCo, estão sendo executadas até o momento da escrita desta tese e outros dados como amostras de sangue, dados do *Brain Explorer*, exames de imagem serão adicionados à base de dados.

2.4 Algoritmos e métodos de AM

Nos últimos anos, algoritmos de AM possibilitaram o desenvolvimento de modelos com potencial para otimizar o diagnóstico da depressão. Nesta seção introduzimos os algoritmos e técnicas de aprendizado de máquina utilizados neste estudo.

2.4.1 Algoritmos de Aprendizado Clássico

Nós adotamos cinco algoritmos de aprendizado clássico para investigar a depressão na adolescência: *Random Forest* (RF), *Extra trees* (ET), *Gradient Boosting* (GB), *Catboost* (CB) e Máquinas de Vetores de Suporte (SVM - do inglês *Support Vector Machine*).

RF (BREIMAN, 2001) e ET (GEURTS; ERNST; WEHENKEL, 2006) são algoritmos de aprendizado *ensemble* que combinam múltiplas árvores de decisão individuais para melhorar o desempenho do modelo e realizar previsões mais precisas. Eles podem ser adotados em tarefas de classificação e regressão.

O RF utiliza um método (*Bagging*) para escolher aleatoriamente subconjuntos de variáveis e de dados de treinamento do modelo. Em seguida, árvores de decisão independentes são construídas para cada subconjunto aleatorizado. A previsão final do modelo é dada pela votação majoritária dos modelos preditivos (LIAW; WIENER et al., 2002).

O RF foi utilizado para classificar usuários com depressão por meio de fotos postadas no Instagram (REECE; DANFORTH, 2017), e para classificar a depressão através de tuítes (CHEN et al., 2018). Adicionalmente, a solução vencedora do AVEC 2017 utilizou o RF como modelo de regressão (RINGEVAL et al., 2017). Ele também é um algoritmo bastante usado para a classificação de depressão em dados multimodais.

O ET é extensão do algoritmo RF. A principal diferença entre eles está no nível de aleatoriedade introduzido pelos algoritmos na construção das árvores de decisão. O ET é considerado mais aleatório, porque os subconjuntos de dados são formados aleatoriamente para cada árvore de decisão adicionada ao modelo.

O GB (FRIEDMAN, 2001) e CB (PROKHORENKOVA et al., 2018) são algoritmos *ensemble* baseados no método *Boosting*. Estes algoritmos constroem modelos sequenciais, normalmente árvores de decisão, onde a cada nova iteração, o modelo subsequente é treinado para corrigir os erros do modelo anterior. Ambos podem ser usados em tarefas de classificação e regressão.

O GB combina um conjunto de modelos fracos em sequência, para criar um modelo mais forte. Ele usa uma técnica de otimização baseada no cálculo do gradiente para minimizar a função de perda do modelo anterior, encontrando os melhores pesos e parâmetros de cada novo modelo. O CB é uma variação do GB de código aberto otimizada para suportar variáveis categóricas (FRIEDMAN, 2001; PROKHORENKOVA et al., 2018).

O SVM (CORTES; VAPNIK, 1995) é um algoritmo bastante utilizado em trabalhos de detecção de depressão devido a sua versatilidade e capacidade de predição. Ele estabelece como resultado o hiperplano que melhor separa os pontos de dados de diferentes classes, para maximizar a distância entre os pontos de dados próximos ao hiperplano.

O algoritmo SVM é muito empregado em predições binárias como, por exemplo, para classificar pessoas com depressão e sem depressão (LIU et al., 2020), mas também pode ser aplicado em tarefas de regressão (JADHAV; SUGANDHI, 2018).

O SVM foi usado para identificar a depressão em textos do Twitter e Facebook (DE CHOUDHURY; COUNTS; HORVITZ, 2013; SHUAI et al., 2017), em dados acústicos (JIANG et al., 2017) e em entrevistas em vídeo (ANIS et al., 2018). Ele também foi adotado como modelo de regressão para identificar a depressão em dados multimodais, especificamente, áudio, vídeo e transcrições de entrevistas da base DAIC-WOZ (STEPANOV et al., 2018).

2.4.2 Algoritmos de Aprendizado Profundo

Neste estudo, também adotamos dois algoritmos de aprendizado profundo bastante populares, uma Rede Neural Profunda (DNN - do inglês *Deep Neural Network*) e uma Rede Neural Convolucional de 1 Dimensão (1DCNN - do inglês *One dimensional Convolutional Neural Network*). Ambas podem ser usadas em tarefas de classificação e regressão e manipular diversos tipos de dados.

2.4.2.1 Redes Neurais Profundas

DNN são uma arquitetura de rede neural artificial formada por múltiplas camadas intermediárias entre a camada de entrada e a camada de saída da rede. Essas camadas são empilhadas e formadas por várias unidades de processamento (neurônios) interconectados (LIU et al., 2017).

Elas são treinadas usando o algoritmo *backpropagation*, que ajusta os pesos das ligações entre neurônios durante o treinamento para minimizar a diferença entre os valores reais e a saída prevista pelo modelo. Assim, cada camada da rede recebe e processa informações da camada anterior, até a previsão final na camada de saída.

2.4.2.2 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN - *Convolutional Neural Networks*) representam uma classe importante de arquiteturas de aprendizado profundo projetadas para processamento de dados com estrutura de grade, devido a suas camadas de convolução (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016).

CNNs usam operações de convolução para criar mapas de características locais em dados estruturados de maneira espacial, como imagens e áudio (LECUN; BENGIO; HINTON, 2015; JADHAV; SUGANDHI, 2018). Basicamente, elas aplicam filtros (*kernels*), para realizar operações de convolução espacial em regiões específicas dos dados de entrada, reduzindo sua dimensionalidade (*pooling*), mas mantendo características locais importantes dos dados processados em várias camadas da rede (LE-

CUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016). Por esse motivo, são amplamente utilizadas no processamento de dados multidimensionais, tarefas de visão computacional e análise de imagens.

CNNs foram adotadas para prever a depressão em vídeos por meio de tarefas de classificação (PAMPOUCHIDOU et al., 2017); ou trabalhos de regressão (ZHOU et al., 2018). Também podem ser encontrados modelos preditivos de depressão baseados em CNN que usaram dados de áudio (DUBAGUNTA; VLASENKO; DOSS, 2019) e multimodais (LAM; DONGYAN; LIN, 2019; YANG et al., 2017). Além disso, redes 1DCNN são uma variação das redes convolucionais projetadas para processar dados unidimensionais e têm sido boas alternativas para classificar emoções em áudios, sequências de texto e dados multimodais.

2.4.3 Métodos de validação cruzada

O processo de validação cruzada envolve a subdivisão dos dados de treinamento em várias partições ou dobras, alternando as dobras usadas para a validação em cada iteração.

A validação cruzada fornece estimativas mais precisas do desempenho de modelos de AM, uma vez que os testa em diferentes dados de validação, garantindo que o desempenho geral do modelo seja uma média representativa de sua capacidade de generalização para novos dados não vistos. Técnicas de validação cruzada são especialmente úteis em conjuntos de dados pequenos, reduzindo problemas de superajuste e subajuste dos modelos preditivos.

2.4.3.1 Validação cruzada estratificada

O método de validação cruzada estratificada é bastante utilizado para problemas de classificação e em conjunto de dados com uma distribuição de classes desequilibrada. Esse método equilibra a distribuição das amostras em cada dobra da validação cruzada para que elas mantenham a mesma proporção do conjunto de treinamento, assegurando que cada dobra seja representativa da distribuição geral (BERRAR et al., 2019).

2.4.3.2 Validação cruzada Leave One Group Out

O método de validação cruzada deixar um grupo de fora (LOGO - do inglês *Leave One Group Out*) é utilizado para avaliar o desempenho de modelos de AM, quando as amostras dos conjuntos de dados podem ser organizadas em grupos distintos.

Na validação cruzada LOGO, a cada iteração, um único grupo de dados é selecionado para teste, enquanto os outros grupos são usados para treinar o modelo preditivo. Esse processo é repetido até que cada grupo de amostras seja usado como conjunto de validação durante o treinamento do modelo (ARLOT; CELISSE, 2010).

A LOGO pode ser útil em estudos longitudinais ou quando há agrupamentos naturais nos dados. Esse método oferece uma estimativa de desempenho do modelo menos enviesada, uma vez que todos os grupos são utilizados tanto para treinamento quanto para teste do desempenho do modelo. No entanto, ele é bastante sensível à qualidade das amostras do grupo e pode ser computacionalmente mais caro quando o número de grupos é grande (ARLOT; CELISSE, 2010).

2.4.3.3 Validação cruzada de séries temporais

O método de validação cruzada *TimeSeriesSplit* (PEDREGOSA et al., 2011) é especializado em séries temporais. Ele é bastante utilizado em pesquisas que priorizam as dependências temporais dos dados para avaliar o desempenho de modelos de AM como, por exemplo, previsões financeiras e previsão do tempo.

Na validação cruzada *TimeSeriesSplit*, as amostras são organizadas sequencialmente considerando a ordem temporal dos dados. Durante o processo de validação cruzada, cada iteração adiciona mais dados históricos ao conjunto de treinamento, evitando o vazamento de informações entre amostras do futuro para o passado (PEDREGOSA et al., 2011). Porém, uma série muito extensa pode tornar o treinamento dos modelos de AM, mais caro, computacionalmente, se comparado a outros métodos de validação cruzada.

2.4.4 Técnicas de otimização de hiperparâmetros

Neste estudo, adotamos as técnicas de otimização de hiperparâmetros, busca aleatória e otimização bayesiana, para explorar o espaço de hiperparâmetros dos modelos preditivos.

2.4.4.1 Busca aleatória

O método de busca aleatória para otimização de hiperparâmetros *RandomizedSearchCV* (BERGSTRA; BENGIO, 2012) oferece um meio eficiente para explorar o espaço de hiperparâmetros de um modelo de AM.

O *RandomizedSearchCV* seleciona aleatoriamente valores do espaço de hiperparâmetros sob uma distribuição de probabilidade (amostragem) para encontrar conjunto ótimo de hiperparâmetros durante o treinamento dos modelos, permitindo flexibilizar o número de iterações e o tempo dedicado ao processo de otimização.

A busca aleatória é usada especialmente quando o espaço de hiperparâmetros a ser explorado é grande e mal compreendido, permitindo que a busca seja mais rápida e menos custosa em termos computacionais, se comparado a outros métodos como a busca em grade (*Grid Search*) pois não precisa explorar todas as combinações do espaço de hiperparâmetros.

2.4.4.2 Otimização bayesiana

A otimização bayesiana é um método avançado de busca de hiperparâmetros em modelos de AM baseada em técnicas probabilísticas e de modelagem estatística (AKIBA et al., 2019).

O processo de otimização bayesiana utiliza um método probabilístico que estima a relação entre os hiperparâmetros e a métrica de desempenho do modelo. Este método é atualizado a cada nova interação durante o treinamento. Assim, o conhecimento gerado durante o processo de otimização das iterações anteriores é utilizado para ajustar os hiperparâmetros nas próximas iterações (BERGSTRA et al., 2011; AKIBA et al., 2019).

A otimização Bayesiana bastante útil quando o espaço de busca é grande e complexo, ao permitir explorar de maneira eficiente o conjunto de hiperparâmetros de um modelo de AM, priorizando regiões mais promissoras do espaço de busca durante o processo de treinamento, (BERGSTRA et al., 2011).

2.4.5 Técnicas de sobreamostragem de dados

Conjuntos de dados desequilibrados podem introduzir vieses ao treinamento dos modelos de AM, pois os algoritmos tendem a superestimar a classificação das amostras majoritárias. Neste contexto, métodos de sobreamostragem ou subamostragem de dados são amplamente utilizados em conjuntos de dados desequilibrados, sobretudo quando algumas classes são bastante reduzidas. Existem vários métodos de sobreamostragem de dados, em seguida, apresentamos alguns métodos adotados neste estudo.

2.4.5.1 SMOTE

A Técnica de Sobreamostragem Sintética de Minorias (SMOTE - do Inglês *Synthetic Minority Over-sampling Technique*) foi proposto por Chawla et al. (2002) para minimizar o desequilíbrio de classes dos conjuntos de dados em problemas de classificação. O SMOTE foi projetado para criar novas amostras sintéticas da classe minoritária por meio de um processo de interpolação baseado na distância entre as essas amostras e seus vizinhos.

Além disso, o *Borderline SMOTE* (HAN; WANG; MAO, 2005) é uma variante do algoritmo SMOTE original, capaz de identificar amostras da classe minoritárias que estão próximas da fronteira entre as classes usadas para gerar novas amostras sintéticas da classe minoritária e adicioná-las ao conjunto de treinamento.

A principal diferença entre as duas versões do SMOTE é que o SMOTE original gera novas amostras sintéticas da classe minoritária indiscriminadamente, enquanto o *Borderline SMOTE* gera novas amostras sintéticas considerando apenas às amostras que estão próximas à fronteira entre as classes que teoricamente são mais difíceis de

classificar. O SMOTE segue sendo constantemente atualizado, outras variações do algoritmo podem ser consultadas em Kovács (2019).

2.4.5.2 RandomOverSampler

O *RandomOverSampler* (LEMAITRE; NOGUEIRA; ARIDAS, 2017) também é um método de sobreamostragem usado para equilibrar o conjunto de dados em tarefas de classificação. Esse método é mais simples que os métodos abordados anteriormente.

Basicamente, o *RandomOverSampler* aumenta o número de instâncias da classe minoritária através da replicação aleatória das amostras existentes, até que haja um equilíbrio mais próximo entre as classes. Em nosso estudo, adotamos uma solução mista usando *Borderline SMOTE* e *RandomOverSampler* para balancear nossos conjuntos de dados.

2.4.6 Métricas de avaliação

2.4.6.1 Métricas básicas

Os valores para classes positivas e negativas, definem as métricas de classificação. Uma classificação positiva (P) indica o número de adolescentes diagnosticados com depressão, e uma classificação negativa (N) representa os adolescentes identificados sem sintomas de depressão. Onde:

- Verdadeiro Positivo (TP - do inglês *True Positive*): representa os adolescentes deprimidos classificados corretamente;
- Falso Positivo (FP - do inglês *False Positive*): representa os adolescentes sem depressão que o modelo classificou como deprimidos;
- Verdadeiro Negativo (TN - do inglês *True Negative*): representa os adolescentes não deprimidos classificados corretamente;
- Falso Negativo (FN - do inglês *False Negative*): representa os adolescentes deprimidos que o modelo classificou como não deprimidos;

2.4.6.2 Métricas derivadas

A Acurácia corresponde a taxa de diagnósticos corretos; representa a capacidade geral de classificação do modelo. É definida como:

$$Acurácia = \frac{TP + TN}{TP + FP + TN + FN}; \tag{1}$$

A Sensibilidade, Revocação, ou Taxa de Verdadeiro Positivo (TPR - do inglês *True Positive Rate*) calcula a proporção de pacientes com depressão classificados corretamente. Ela é dada por:

O Revocação calcula as detecções positivas identificadas corretamente e é dada por:

$$Sensibilidade = Recall = TPR = \frac{TP}{TP + FN}; \tag{2}$$

A Especificidade, Taxa de Verdadeiro Negativo (TNR - do inglês *True Negative Rate*), calcula a proporção de pacientes sem depressão classificados corretamente. Ela é dada por:

$$Especificidade = TNR = \frac{TN}{TN + FP}; \tag{3}$$

A Precisão ou Valor Preditivo Positivo (PPV - do inglês *Positive Predictive Value*) calcula a probabilidade de um paciente classificado como positivo ser da classe deprimida. Ela é dada por:

$$Precisão = PPV = \frac{TP}{TP + FP}; \tag{4}$$

A Medida-F1 é dada pela média harmônica entre Revocação e Precisão.

$$Medida\ F1 = \frac{2 \cdot Precisão \cdot Recall}{Precisão + Recall}; \tag{5}$$

A Acurácia Balanceada binária mede a capacidade do modelo para discriminar as classes positivas e negativas em conjuntos de dados desbalanceados. Ela é calculada pela média aritmética da Sensibilidade e Especificidade:

$$Acurácia\ Balanceada = \frac{Sensibilidade + Especificidade}{2}; \tag{6}$$

As métricas mais comumente adotadas em tarefas de regressão para prever a depressão são Erro Médio Absoluto (MAE) e a Raiz quadrada do erro médio (RMSE).

O MAE calcula a média das diferenças absolutas entre as previsões e os valores reais, dada pela fórmula:

$$MAE = (\frac{1}{n}) \sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{7}$$

O RMSE calcula a raiz quadrada da média dos erros quadráticos entre as previsões e os valores reais. Ela é dada por:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}; \tag{8}$$

Assim, MAE e o RMSE representam o valor do desvio médio entre o que é predito e a verdade básica. Ambas consideram a magnitude do erro, não a sua direcionalidade. A principal diferença entre elas é que o RMSE penaliza mais os desvios, atribuindo

pesos maiores aos erros do modelo, por ser uma função quadrática. Em nosso estudo, essas medidas foram utilizadas em conjunto para avaliar os modelos preditivos.

2.4.7 *Shapley Additive exPlanations* (SHAP)

Neste estudo, nós adotamos três gráficos de importância derivados do algoritmo SHAP (LUNDBERG; LEE, 2017) para investigar quais atributos, ou uma combinação entre eles, contribuem para otimizar o desempenho dos modelos preditivos de depressão.

O Gráfico de importância global (Figura 46) calcula a contribuição dos atributos para prever as pontuações de depressão binária do sMFQ. Ela é dada pelo valor absoluto médio para cada atributo em todas as amostras do conjunto de testes. Eles são organizados em ordem decrescente de importância, de modo que atributos com valores de SHAP mais altos no conjunto de teste, contribuem mais para classificar as amostras (LUNDBERG; LEE, 2017).

O gráfico de direcionalidade (Figura 47) estima o impacto dos principais atributos na saída do modelo preditivo. No eixo y, estão listados os atributos mais importantes para a previsão do modelo, organizados da mesma forma que o gráfico de importância global. No eixo x, os valores de SHAP estão separados por uma linha central.

Valores de SHAP positivos (à direita da linha central) para um determinado atributo indicam a influência deste para a previsão da classe deprimida. Ao mesmo tempo, valores de SHAP negativos (à esquerda da linha central) indicam o impacto do atributo na previsão da classe não deprimida.

A cor do ponto representa o valor do atributo, os pontos azuis para valores mais baixos e pontos vermelhos para valores mais altos. Eles são distribuídos horizontalmente no gráfico, cada ponto representa uma previsão das amostras do conjunto de testes e como ela é afetada por aquele atributo.

Para a interpretação dos resultados, nós consideramos que, quanto mais dispersos horizontalmente são os pontos, ou havendo concentração em locais afastados da linha central, maior a influência do atributo na previsão do modelo.

Além disso, nós adicionamos um gráfico de direcionalidade para cada atributo (Figura 48). Esse gráfico foi benéfico para visualizar a capacidade de cada atributo individualmente para discriminar amostras deprimidas das amostras não deprimidas durante os experimentos.

3 TRABALHOS RELACIONADOS

Este Capítulo revisa os avanços mais recentes alcançados pelos modelos de AM supervisionados aplicados ao contexto da depressão. Os estudos selecionados buscam analisar o comportamento depressivo mediante dados multimodais coletados via *smartphones* ou sensores vestíveis exclusivamente relacionados à depressão.

3.1 Seleção dos trabalhos relacionados

Nós coletamos artigos em periódicos e conferências internacionais das bases de dados *Association for Computing Machinery* (ACM); *Institute of Electrical and Electronics Engineers* (IEEE), *Science Direct* e *PubMed*. A busca e seleção dos trabalhos ocorreu entre os meses de agosto e dezembro de 2022. Os operadores *AND* e *OR* foram adotados para combinar os descritores *depression*, *smartphone*, *machine learning* e *deep learning*. Foram considerados artigos publicados entre 2019 e 2022. Eles foram selecionados com base nos seguintes critérios de inclusão e exclusão:

- Critérios de inclusão:
  - Artigos baseados em AM aplicados ao contexto da depressão;
  - Artigos que usam dados extraídos de *smartphones* ou sensores vestíveis;
- Critérios de exclusão:
  - Artigos que não estão relacionados exclusivamente com a depressão;
  - Artigos que não usam dados multimodais
  - Artigos que não realizam experimentos

A Figura 3 apresenta uma visão geral de estudos baseados em AM supervisionados aplicados ao contexto da depressão. Ela descreve um pipeline básico para o desenvolvimento de modelos de predição de depressão estabelecidas na literatura.

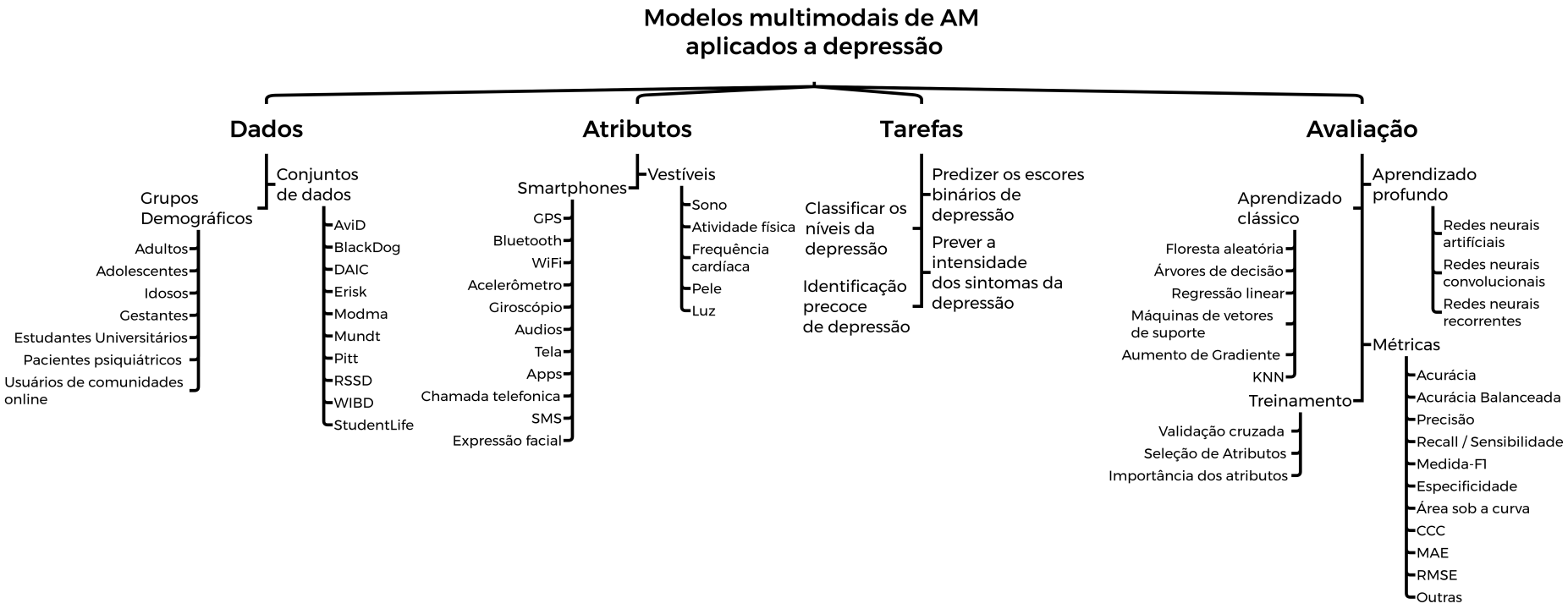


Figura 3 – Visão Geral das estratégias de detecção de depressão. Fonte: Autoria própria

A Tabela 5 apresenta o número de artigos avaliados durante a etapa de seleção em cada base de dados. No total, encontramos 1.426 artigos e, após verificação dos critérios de inclusão e exclusão, selecionamos 13 estudos para desenvolver este capítulo.

Tabela 5 – Detalhamento do processo de seleção dos trabalhos relacionados.

Bases	Encontrados	Incluídos	Excluídos	Selecionados
AMC	856	11	9	2
IEEE	17	4	2	2
Pubmed	64	22	16	6
Science direct	489	13	10	3

Metodologicamente, este capítulo está organizado para dar ênfase às diversas fontes de dados, formas de seleção de atributos, arquiteturas disponíveis e métricas usadas para avaliar modelos de AM em tarefas distintas de detecção de depressão. Mais detalhes sobre os trabalhos selecionados podem ser consultados no Apêndice A.

3.2 Análise dos trabalhos relacionados

Os modelos de AM multimodais analisados neste estudo, investigaram dados coletados de *smartphones* e sensores vestíveis como marcadores digitais da depressão. A Tabela 6 apresenta informações sobre o perfil sociodemográfico dos participantes, número de amostras, período de coleta e melhores resultados obtidos nos estudos que compõem este capítulo.

Tabela 6 – Caracterização dos estudos de detecção de depressão. O nome de algumas colunas contendo o * foram abreviados: Participantes (N), Intervalo (I), Média (M) semanas (s) e dias (d).

Referência	Origem	Grupos	N*	Idade* (I)	Idade* (M)	Sensores	Apps / dispositivos*	Coleta	Validação	Melhores resultados
Masud et al. (2020)	Bangladesh	Adultos	33	18	24	*Smartphone*	*Data Collector*	11s	PHQ-9	Previu indivíduos com sintomas graves de depressão com Acc de 87,2%
Tazawa et al. (2020)	Japão	Idosos	86	20	59	Vestíveis	*Silmee W11*	7d	HDRS-17	Acc de 74% e 76%, para o conjunto de dados de 3 e 7 dias, respectivamente.
Pedrelli et al. (2020)	EUA	Adultos	31	19-73	33	*Smartphone* e vestíveis	*Empathic E4*, MovisensXS	8s	HDRS-17	Valores de MAE entre 3,88 e 4,74 usando somente atributos extraídos via *smartphones*
Asare et al. (2021)	EUA	Adultos	629	18-65	-	*Smartphone*	Carat	22d	PHQ-8	Taxas de 94,03% de Acc e 95,27% de F1 para prever da classe deprimida.

Referência	Origem	Grupos	N*	Idade* (I)	Idade* (M)	Sensores	Apps / dispositivos*	Coleta	Validação	Melhores resultados
Chikersal et al. (2021)	EUA	Estudantes universitários	138	18+	-	*Smartphone* e vestíveis	*AWARE*, *Fitbit Flex2*	16s	BDI-II	prever a depressão pós-semestre e alteração da depressão durante o semestre com Acc de 85,7% e 85,4%, respectivamente.
Nguyen et al. (2021)	EUA	Estudantes universitários	48	18+	-	*Smartphone*	-	10s	PHQ-9	Classificar os estudantes em 3 grupos de pontuação do PHQ-9 com Acc de 94,7%
Hong et al. (2022)	Coreia do Sul	Pacientes de clínicas psiquiátricas	106	18+	-	*Smartphone*	*Mental Health Protector*	2s	PHQ-9 e CESD-R	prever os escores binários do PHQ-9 e CESD-R com Acc de 74,07% e 76,92%
Choudhary et al. (2022)	Coreia do Sul	Adultos	558	18+	-	*Smartphone*	*Behavior Research*	10,7d	PHQ-9	Prever os escores binários e 3 níveis de sintomas do PHQ-9, com Acurácia de 87% e 78%, respectivamente.

Referência	Origem	Grupos	N*	Idade* (I)	Idade* (M)	Sensores	Apps / dispositivos*	Coleta	Validação	Melhores resultados
Zhong et al. (2022)	Suécia	Gestantes suecas	915	18+		*Smartphone*	*Mob2b*	42s	EPDS	Taxas de B.Acc de 75% e uma área sob a curva de 82%
Mullick et al. (2022)	EUA	Adolescentes	37	12-17	15,5	*Smartphone* e vestíveis	*AWARE*, Fitbit Inspire HR	24s	PHQ-9	Prever o escore de depressão e a mudança semanal no nível de depressão com valores de MAE de 2,39 e 3,21
Yan; Tu; Wen (2022)	China	Adultos	263	20-60	-	*Smartphone* e vestíveis	Ferramenta proprietária	5d	DASS-21	Acc de 83%, F1 de 79% e AUC de 80%
Asare et al. (2022)	EUA	Usuários de comunidades *on-line*	54	18+	-	*Smartphone* e vestíveis	*Oura Ring*	4s	DASS-21	Prever as classes deprimida e não deprimida com valores de Acc de 81,43%, AUC de 82,31% e F1-macro de 73,34%.
Ware et al. (2022)	EUA	Estudantes universitários	59	18-25		*Smartphone*	*LifeRhythm*	2s	HDRS	Acc de 80% na tarefa de classificação binária de depressão em estudantes universitários.

Os estudos apresentados neste capítulo foram realizados com maior frequência nos Estados Unidos, mas também foram conduzidos em outros países como Bangladesh, Japão, Coreia do Sul, Suécia e China. A maioria das pesquisas (85,7%) priorizou as análise de depressão em participantes com idade acima dos 18 anos e metade (50%) dos estudos analisados envolveram participantes que falam o idioma inglês.

A escassez de dados de participantes deprimidos é um fator limitante para o desenvolvimento de estudos sobre a depressão. O número de amostras analisadas nos estudos mostrou-se bastante reduzido, dado que 50% dos estudos eram formados por até 100 participantes, 21,5% dos trabalhos tiveram entre 100 e 200 participantes e 28,5% tiveram acima de 200 participantes envolvidos na pesquisa.

O período de coleta de dados dos participantes se mostrou bastante heterogêneo, variando entre 5 dias até 42 semanas. Além disso, a quantidade de participantes também varia significativamente entre os estudos (entre 33 e 915 participantes). Essa variabilidade temporal e entre os participantes pode comprometer a capacidade de generalização dos modelos e a relevância estatística da pesquisa.

Os estudos investigaram diferentes grupos demográficos de várias partes do mundo. Especificamente, esses estudos avaliaram a depressão em mulheres grávidas suecas que estavam entre o 1.º e o 2.º trimestre da gestação (ZHONG et al., 2022), em estudantes universitários Nguyen et al. (2021); Chikersal et al. (2021); Ware et al. (2022), em usuários de comunidades online (ASARE et al., 2021), em pacientes psiquiátricos (HONG et al., 2022), idosos (TAZAWA et al., 2020) e em adolescentes (MULLICK et al., 2022). Outros estudos investigaram a depressão em indivíduos sem especificarem um grupo demográfico (MASUD et al., 2020; PEDRELLI et al., 2020; ASARE et al., 2021; CHOUDHARY et al., 2022; NGUYEN et al., 2021; YAN; TU; WEN, 2022).

Os modelos preditivos desenvolvidos conseguem manipular dados de diversos tipos. Apenas um artigo usou apenas dados de sensores vestíveis para prever a depressão. 57,1% dos artigos usaram atributos extraídos apenas de *smartphones* e 35,7% dos artigos compararam simultaneamente dados de *smartphones* e sensores vestíveis. Além disso, os estudos de Asare et al. (2021); Chikersal et al. (2021); Asare et al. (2022) obtiveram ganhos de desempenho ao adicionar dados sociodemográficos ao modelo preditivo.

Os resultados obtidos pelos estudos variam consideravelmente, no entanto, os modelos desenvolvidos demonstraram alta capacidade preditiva. Eles usaram em sua maioria medidas de classificação como Acurácia, Acurácia Balanceada, AUC, Medida-F1 e medidas de regressão como MAE. A Tabela 7 resume as principais tarefas propostas, atributos utilizados, algoritmos de classificação e principais resultados desses estudos.

Tabela 7 – Principais avanços de modelos multimodais de AM para diagnosticar a depressão. Algoritmos: SVM; CB; GB; K-Nearest Neighbors (KNN); Multi-Layer Perceptron (MLP); RF; Extreme Gradient Boosting (XGB); Logistic Regression (LR); Decision Tree (DT); ET; DNN ; 1DCNN.

Referência	Tarefa	Atributos	Modalidades	Método	Achados
Masud et al. (2020)	Classificar níveis de depressão (ausente, moderada e grave) com base no PHQ-9	12	Atividade e mobilidade	SVM, KNN, MLP, validação cruzada de dez vezes e algoritmos de seleção de atributos	A depressão grave foi associada com valores baixos de variação de distância, passos, entropia normalizada e exercícios e com mais tempo de descanso, permanência em casa e uso do telefone deitado
Tazawa et al. (2020)	prever escores binários de depressão com base em dados coletados em 3 e 7 dias antes da avaliação clínica	63	Demográficos, passos, energia, movimento corporal, sono, frequência cardíaca, temperatura e exposição à luz ultravioleta	SVM, XGBoost, e RF e validação cruzada de dez vezes.	A temperatura da pele e o tempo de sono foram os atributos mais importantes para prever a depressão em idosos.
Pedrelli et al. (2020)	Prever a intensidade dos sintomas da depressão com base no HDRS-17.	887	Uso do *smartphone, uso de apps*, localização, chamadas telefônicas, SMS, tela, atividade eletrotérmica, pele, frequência cardíaca, movimento e sono.	*Adaboost*, RF, método *Ensemble* e validação cruzada de dez vezes com seleção de atributos (Boruta e RFE)	O modelo que usou apenas atributos dos *smartphones* teve o melhor desempenho para estimar mudanças na intensidade dos sintomas da depressão.

Referência	Tarefa	Atributos	Modalidades	Método	Achados
Asare et al. (2021)	prever escores binários de depressão com base no PHQ-8	22	*Status* da tela, conectividade com a internet e logs de uso de aplicativos e demográficos.	RF, SVM, XGB, KNN e LR, validação cruzada estratificada e aninhada com importância dos atributos	Os atributos de conectividade de tela e conectividade com a internet foram considerados mais importantes para prever a depressão.
Chikersal et al. (2021)	prever escores binários de depressão em universitários pós-semestre e a alteração da depressão no decorrer do semestre	62311	Uso do telefone, chamadas telefônicas, localização, bluetooth, mapa do campus, passos e sono.	LR e *Gradient Bosting* com validação cruzada de exclusão	Dados de *Bluetooth*, Chamada telefônica, Uso do telefone e Passos foram os mais importantes para detectar a depressão pós-semestre, enquanto os dados de *Bluetooth*, Mapa do campus, Uso do telefone e Sono foram mais importantes para prever a alteração da depressão no decorrer do semestre
Nguyen et al. (2021)	Classificar os níveis de depressão (baixo médio e alto) do PHQ-9 em universitários	33	Atividade, luz, áudios, conversação e bloqueio de telefone	RF, SVM, XGB, KNN e LR com método de validação cruzada de 10 vezes	Os universitários com baixa pontuação PHQ-9 foram associados a mais atividades de caminhadas e conversas realizadas pelo *smartphone*

Referência	Tarefa	Atributos	Modalidades	Método	Achados
Mullick et al. (2022)	prever escores binários e mudanças nos níveis de depressão em adolescentes com base no PHQ-9	66	Chamadas e conversas telefônicas, localização, *status* de tela, Wi-Fi, frequência cardíaca, sono e passos.	RF, AdaBoost e XGB, varios métodos de validação cruzada com importância dos atributos	Os atributos baseados em tela, chamada telefônica e localização foram considerados os mais importantes preditores da depressão em adolescentes.
Choudhary et al. (2022)	prever escores binários e níveis de depressão (ausente, moderada e grave) do PHQ-9	37	Tempo de uso do celular; frequência de uso dos aplicativos; giroscópio.	*Regression splines*, RF, XGB e SVM e validação cruzada de 15 vezes	O tempo médio da sessão e o uso de aplicativos de interação social (categoria 1) foram maiores em participantes com depressão grave.
Zhong et al. (2022)	Identificação precoce de depressão em mulheres gestantes suecas	-	Sociodemográficos, saúde psicológica, saúde geral, comportamento, social e personalidade	LR, SVM, KNN, XGB e MLP com validação cruzada estratificada de 5 vezes	O conjunto de dados de saúde psicológica teve maior contribuição para detectar a depressão precoce em gestantes.
Hong et al. (2022)	prever escores binários de depressão de pacientes de clínicas psiquiátricas com base no PHQ-9 e CESD-R	33	Sono, atividade, e expressão facial	RF	Os atributos de expressão facial foram mais importantes para prever os escores binários do PHQ-9, mas os modelos multimodais superaram o desempenho dos modelos unimodais.

Referência	Tarefa	Atributos	Modalidades	Método	Achados
Yan; Tu; Wen (2022)	prever escores binários de depressão do DASS-21 em adultos chineses	10	Uso do telefone, atividade física, dieta, sono e localização	1DCNN, SVM, LR, RF, MLP, DT	Os atributos mais importantes conforme o método de discretização foram número de vezes que usou o telefone, tempo de duração do uso do telefone e calorias ingeridas.
Asare et al. (2022)	prever escores binários de depressão do DASS-21 em universitários	28	Atividade, mobilidade, uso do telefone, humor e sono	RF, SVM, XGB, KNN e LR com método de validação cruzada aninhada	Os atributos de uso do telefone, mobilidade e sono foram considerados os mais importantes para prever a depressão binária em universitários
Ware et al. (2022)	prever escores binários de depressão em universitários	59	SMS e de chamadas telefônicas.	RF, SVM e XGBoost e método de validação deixar um de fora	Atributos de SMS e chamadas telefônicas combinados melhoraram o desempenho dos modelos desenvolvidos.

3.2.1 Tarefas de detecção de depressão

Os modelos suportam abordagens multi-atributos, sendo avaliados em diversas tarefas de detecção da depressão com alta capacidade preditiva. Identificamos quatro grupos de tarefas de detecção da depressão: a predição do escore binário de depressão, predição dos níveis de depressão, predição dos sintomas da depressão e identificação precoce da depressão.

A maioria dos estudos adotou a tarefa de predição do escore binário de depressão para identificar indivíduos deprimidos e não deprimidos com base em vários questionários psiquiátricos. Asare et al. (2021) propuseram um modelo de classificação binária da depressão baseado no ponto de corte do PHQ-8. O estudo de Hong et al. (2022) usou dados de *smartphones* para prever a depressão em pacientes de clínicas psiquiátricas coreanas. Os autores usaram as pontuações de corte binárias do PHQ-9 e CESD-R para avaliar o desempenho do modelo preditivo. Em Yan; Tu; Wen (2022), uma rede 1DCNN foi desenvolvida para prever os escores binários de depressão do DASS-21 em adultos chineses usando dados sequenciais coletados via *smartphones* e pulseiras eletrônicas.

Asare et al. (2022) e Ware et al. (2022) propuseram modelos de classificação binária para prever a depressão em estudantes universitários. O primeiro avaliou uma série de dados de *smartphones* e sensores vestíveis como marcadores de depressão, enquanto o segundo usou dados de SMS e chamadas telefônicas para avaliar as relações entre a interação social dos estudantes e a depressão.

Masud et al. (2020) e Nguyen et al. (2021) desenvolveram modelos de AM capazes de prever três níveis de gravidade dos sintomas da depressão com base na pontuação do PHQ-9.

Pedrelli et al. (2020), usaram modelos de AM para prever a gravidade dos sintomas da depressão. Eles adotaram métodos de regressão para identificar a pontuação exata do questionário HDRS preenchido pelos idosos participantes da pesquisa.

Nós identificamos estudos focados na identificação precoce da depressão. Tazawa et al. (2020), avaliaram o desempenho do modelo de AM para prever o escore binário de depressão usando dados de sensores vestíveis coletados 3 e 7 dias antes da avaliação clínica. O estudo de Zhong et al. (2022) propuseram um modelo de AM que utilizou dados coletados no 1º e 2º trimestres de gravidez para identificar depressão em mulheres que preencheram o questionário EPDS entre a 36ª e a 42ª semana de gestação.

Alguns estudos adotaram abordagens multitarefa para detectar a depressão. Por exemplo, Mullick et al. (2022) propuseram um modelo de AM para prever escores binários de depressão e mudanças nos níveis de depressão em adolescentes que completaram o questionário PHQ-9 semanalmente. O modelo que obteve o melhor desempenho usou o método “Semanas Acumuladas” e o classificador XGBoost para

prever o escore de depressão e a mudança semanal no nível de depressão com valores de erro médio MAE de 2,39 e 3,21. Choudhary et al. (2022) também desenvolveram um modelo multitarefa para prever os escores binários e classificar três níveis de depressão (ausente, moderada e grave) do questionário PHQ-9. Os autores agregaram os dados em um período de 24 horas, os modelos usaram classificador *Random Forest* e alcançaram valores de Acurácia de 87% e 78%, respectivamente.

Chikersal et al. (2021) propuseram um modelo de AM que usa dados coletados durante o semestre para prever a ausência ou presença de depressão em estudantes universitários após o final do semestre letivo e para avaliar a mudança na intensidade de sintomas (piorou ou não piorou) durante o semestre. Os modelos com melhor desempenho permitiram detectar a presença da depressão pós-semestre com uma Acurácia de 85,7% e a alteração da depressão durante o semestre com uma Acurácia de 85,4%.

3.2.2 Marcadores digitais da depressão

Os estudos analisados neste capítulo avaliaram atributos extraídos de dispositivos móveis como marcadores digitais da depressão. Sensores de *smartphones* podem oferecer uma variedade de pistas para identificar sintomas da depressão. Os dados mais comumente investigados pelos estudos selecionados neste capítulo estão relacionados à mobilidade, atividade e interação social.

As Figuras 4 e 5, apresentam uma série de atributos extraídos de sensores vestíveis e *smartphones*. Eles podem ser usados individualmente (abordagem unimodal) ou combinados (abordagem multimodal) para o treinamento dos algoritmos.

A mobilidade é medida usando dados de localização capturados passivamente do *smartphone* dos pacientes com depressão. A maioria dos estudos incluídos nesta revisão investigou características de mobilidade como marcadores digitais de depressão. Esses estudos propuseram modelos de AM eficazes para prever a depressão analisando características da mobilidade extraídos das coordenadas de latitude e longitude do GPS, como demostrado nos estudos de Masud et al. (2020); Pedrelli et al. (2020); Yan; Tu; Wen (2022) e (ASARE et al., 2022).

A coleta de dados dos sensores acelerômetro e giroscópio em ambientes domésticos permite monitorar os níveis de atividade de pacientes deprimidos em suas tarefas diárias (PEDRELLI et al., 2020). Masud et al. (2020) usaram dados do acelerômetro e GPS para estimar uma série de atividades físicas humanas. Eles propuseram um estudo focado nas atividades diárias dos participantes e na gravidade da depressão. O estudo indicou que sete atributos tiveram correlação mais forte com a depressão. Os atributos de localização mais importantes foram a variação da distância e entropia normalizada. Os atributos de atividade mais importantes foram repouso, uso do telefone na posição deitada, pisar, se exercitar e ficar em casa.

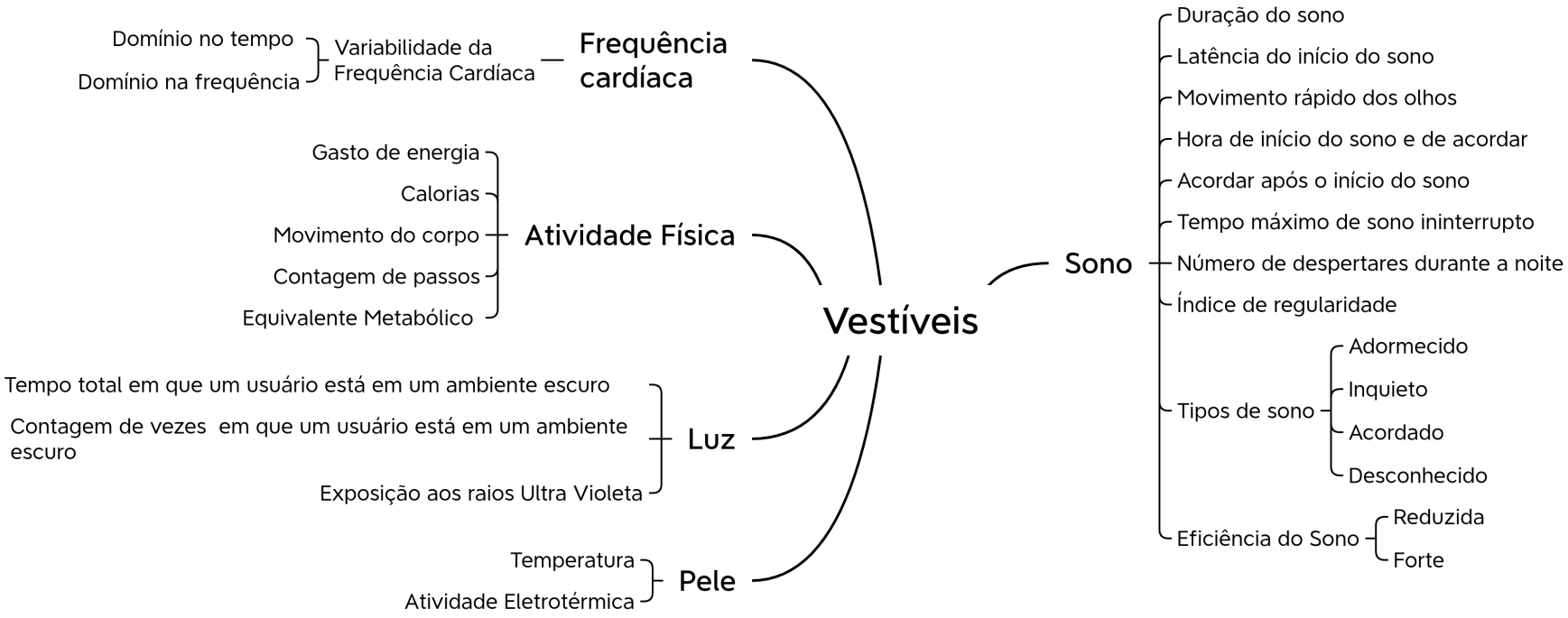


Figura 4 – Marcadores digitais da depressão: Sensores Vestíveis. Fonte: Autoria própria

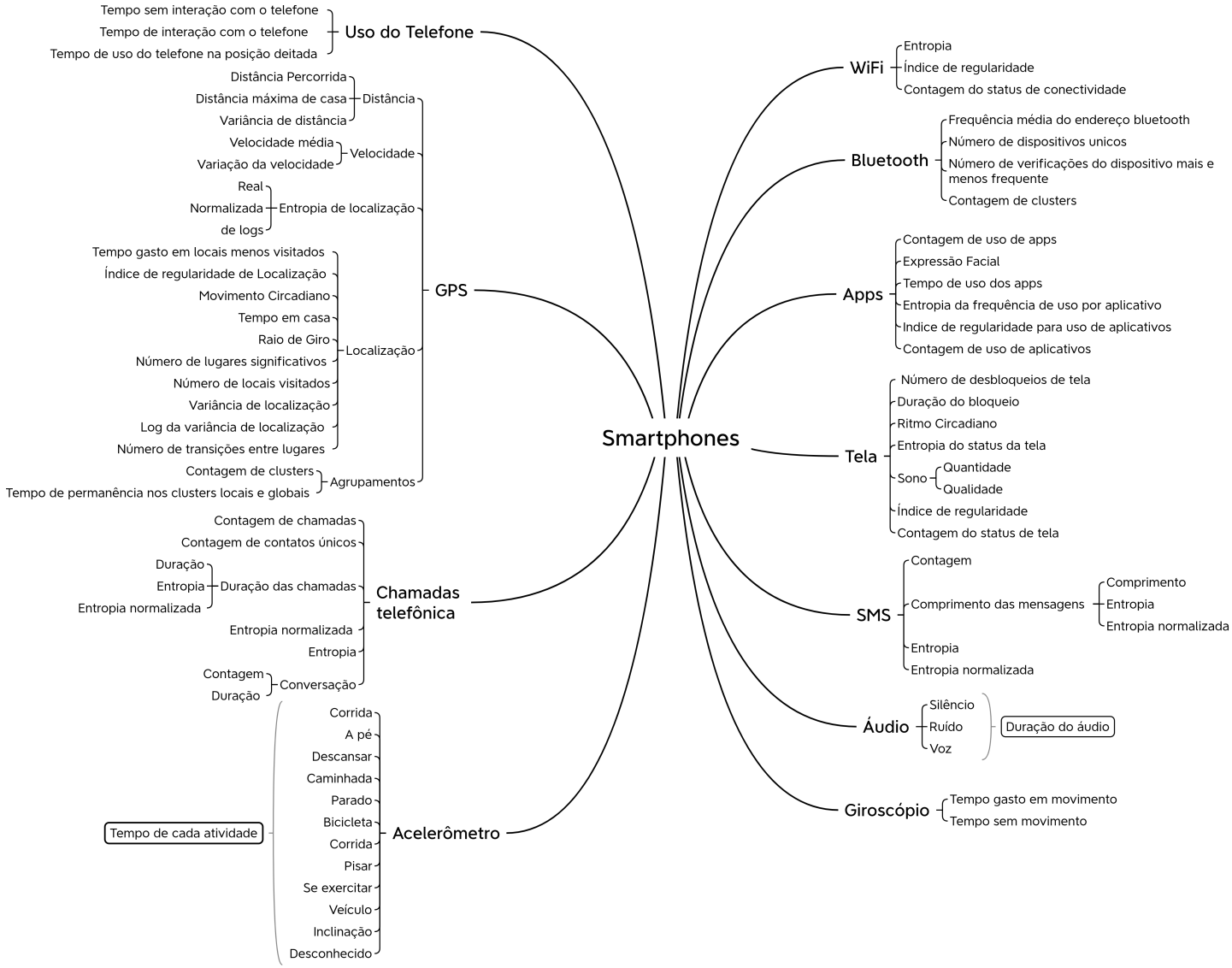


Figura 5 – Marcadores digitais da depressão: Sensores *Smartphones*. Fonte: Autoria própria

Ware et al. (2022) investigaram dados de SMS e chamadas telefônicas para medir a interação social de estudantes universitários. Eles concluíram que os atributos de SMS e chamadas telefônicas combinados melhoram o desempenho dos modelos preditivos. Em Choudhary et al. (2022) dados sobre o uso de aplicativos de categoria 1 (interação social) do aplicativo *Behavidance* foram melhores preditores de depressão grave. Em Yan; Tu; Wen (2022) os dados de uso do telefone foram os mais importantes para prever os escores binários de depressão do DASS-21.

Os estudos analisados neste capítulo adotaram vários sensores vestíveis como marcadores da depressão. Os modelos de detecção de depressão adotaram com mais frequência dados de atividade física e padrões de sono para investigar sintomas de depressão. Em Asare et al. (2022) o tempo total de sono e a eficiência do sono foram considerados os mais importantes para prever a depressão em estudantes universitários. Já Tazawa et al. (2020) concluíram que a temperatura da pele e o tempo de sono foram os atributos mais fortes para o aprendizado do modelo proposto.

Os dados sociodemográficos dos participantes também foram observados na sintomatologia da depressão. Os estudos de Tazawa et al. (2020) e Asare et al. (2021) apontam que incluir características sociodemográficas, especialmente idade, sexo e histórico de saúde do paciente, ao conjunto de dados melhorou o desempenho dos modelos preditivos. Além disso, dados de saúde psicológica tiveram maior contribuição para otimizar o modelo de depressão precoce proposto por Zhong et al. (2022).

3.2.3 Classificadores

Os estudos adotaram algoritmos de aprendizado clássico e aprendizado profundo para desenvolver os modelos de detecção de depressão. Os algoritmos de AM adotados com maior frequência pelos estudos foram *Random Forest*, *Logistic regression*, SVM e XGboost.

Random Forest é um algoritmo bastante usado para classificar a depressão em dados multimodais. Por exemplo, Tazawa et al. (2020) o adotaram para identificar depressão em idosos com dados coletados 3 e 7 dias antes da avaliação clínica. Pedrelli et al. (2020) também adotaram o algoritmo para estimar mudanças na intensidade dos sintomas da depressão mediante uma tarefa de regressão. O modelo proposto por Hong et al. (2022) utilizou o *Random Forest* para prever as pontuações binárias do PHQ-9 com base em dados de sono, atividade e expressão facial.

Como alternativa, o algoritmo *Logistic regression* pode ser usado em tarefas de classificação e regressão. O algoritmo foi avaliado para prever escores binários de depressão do PHQ-8 Asare et al. (2021) e os níveis de depressão de estudantes universitários com base no PHQ-9 Nguyen et al. (2021).

O SVM é um algoritmo de AM amplamente utilizado em trabalhos de detecção de depressão devido a sua versatilidade e capacidade de predição. Ele foi adotado para

classificar os escores binários de depressão (ASARE et al., 2021; CHOUDHARY et al., 2022; WARE et al., 2022) e os níveis de depressão (MASUD et al., 2020; NGUYEN et al., 2021), e em tarefas de identificação precoce da depressão (TAZAWA et al., 2020; ZHONG et al., 2022)

O algoritmo XGboost combina árvores de decisão com o aumento de gradiente. Ele foi utilizado para prever pontuações binárias do PHQ-9 e níveis de depressão em Choudhary et al. (2022); Mullick et al. (2022). O algoritmo foi utilizado em tarefas de identificação precoce de depressão Tazawa et al. (2020), em idosos (CHIKERSAL et al., 2021) e estudantes universitários (ZHONG et al., 2022).

Alguns artigos usaram modelos de detecção de depressão baseados em algoritmos de aprendizado profundo. As Redes Neurais Convolucionais (CNN - *Convolutional Neural Networks*) foram adotadas em Yan; Tu; Wen (2022) para prever os escores binários de depressão em adultos chineses. Além disso, outros modelos de detecção de depressão baseados em *Multi Layer Perceptron* (MLP)e redes neurais profundas Zhong et al. (2022); Yan; Tu; Wen (2022) são encontrados na literatura.

Algumas estratégias de validação cruzada (validação cruzada aninhada, estratificada e de exclusão) e sobreamostragem da classe minoritária (deprimida) foram adotadas durante o treinamento para melhorar o desempenho dos modelos preditivos e minimizar problemas de *overfitting*. Além disso, os estudos frequentemente usam técnicas de seleção e importância de atributos para destacar os atributos mais significativos para previsão dos modelos preditivos de depressão.

3.2.4 Medidas de avaliação

Os modelos de detecção de depressão podem adotar várias métricas para atestar sua confiabilidade. A escolha da métrica depende do problema proposto e/ou do questionário da psiquiatria adotado para validação do conjunto de dados.

Tarefas de classificação são utilizadas quando se avalia, de forma binária, a presença (ou ausência) de depressão; ou categoricamente, os níveis de depressão do questionário. Os métodos de regressão são adotados quando se planeja identificar a pontuação exata do paciente ou dos itens do questionário adotado. A avaliação é perfeita quando um método de classificação atinge o valor máximo de 1, ou um método de regressão atinge um valor próximo a 0.

Os modelos de detecção de depressão foram avaliados com maior frequência em tarefas de classificação considerando as métricas de Acurácia, Precisão, Medida-F1 (F1), medidas de Sensibilidade e Especificidade. Além disso, identificamos medidas mais adequadas para avaliar o desempenho dos modelos na classificação de amostras em conjuntos de dados severamente desbalanceados, tais como Acurácia Balanceada e a área sob a curva (AUC).

As métricas comumente adotadas nos modelos de regressão foram o MAE e o

RMSE. A principal diferença entre eles é que o segundo penaliza mais os desvios ao atribuir maiores pesos aos erros do modelo, por ser uma função quadrática. Na prática, no caso dos modelos preditivos de depressão, esses parâmetros são avaliados em conjunto.

Nós identificamos outras métricas usadas para avaliar os modelos de detecção de depressão como a F1-Macro, Coeficiente de correlação *Pearson* e Coeficiente de Correlação de Concordância (CCC).

3.3 Desafios

Os estudos analisados neste capítulo contribuíram para otimizar o diagnóstico de depressão usando bases de dados bastante heterogêneas e diferentes arquiteturas de AM. Alguns estudos adotaram abordagens multi-atributos e multitarefas para investigar a depressão. Os modelos multimodais frequentemente superaram os modelos unimodais em tarefas de detecção de depressão. Contudo, existem muitas possibilidades de uso de dados de *smartphones* e sensores vestíveis como marcadores digitais da depressão, bem como desafios e limitações a serem superados.

A maioria das pesquisas sobre depressão foi realizada em pacientes adultos. Além disso, os perfis sociodemográficos (idade, sexo, entre outros) e tempo de coleta de dados são, em geral, bastante variados, tornando as amostras dos conjuntos de dados bastante heterogêneas.

A maioria dos estudos investigou a depressão com base em dados coletados em países de língua inglesa e não considera fatores socioeconômicos, culturais e linguísticos que podem dificultar a generalização quando aplicados a um contexto ou cultura diferente.

Outro fator limitante é a aquisição de dados confiáveis sobre a depressão. Redes sociais e plataformas de *crowdsourcing* permitem manipular grandes volumes de dados. No entanto, esses conjuntos de dados foram elaborados de forma pouco estruturada ou padronizada, comprometendo a confiabilidade do estudo, por serem mais suscetíveis a falhas diagnósticas devido ao caráter subjetivo da avaliação. Ainda assim, lidar com dados de pessoas com depressão levanta preocupações sobre a privacidade dos pacientes e está sujeito a sérias restrições éticas e legais.

Além disso, o diagnóstico de depressão é validado por questionários psiquiátricos coletados, em sua maioria, por meio de uma única entrevista ao final do período de coleta de dados. Porém, esse método pode não refletir perfeitamente o quadro do paciente devido às características episódicas da depressão e às flutuações dos sintomas ao longo do tempo.

Grande parte dos trabalhos de detecção de depressão apontam limitações relacionadas à quantidade de amostras disponíveis e a distribuição desequilibrada dos dados

entre as classes (indivíduos deprimidos e não deprimidos). A maioria dos estudos de depressão usou dados de dezenas ou algumas centenas de participantes.

Além disso, dados dos participantes deprimidos estão disponíveis em menor quantidade na maioria dos conjuntos de dados adotados nos estudos. A distribuição desequilibrada das amostras pode introduzir vieses durante o treinamento, que comprometem o desempenho dos modelos preditivos ou a significância estatística da pesquisa. Além disso, fatores socioeconômicos, culturais e linguísticos podem impactar o desempenho de modelos desenvolvidos e são de difícil generalização.

Esta tese realizou um estudo sobre depressão em adolescentes, cujos dados foram coletados através de seus *smartphones*. Ela difere das abordagens anteriores por investigar os sintomas da depressão em adolescentes brasileiros, com faixa etária entre 14 e 16 anos, que estão sendo acompanhados por pesquisadores e profissionais especializados em um estudo longitudinal de três anos.

Foram utilizados métodos de coleta ativa e passiva de dados para compor a base de dados do IDEA-RiSCo. Alguns dados foram gerados mediante questionamentos diretos ao participante (coleta ativa) e também informações de seu estilo de vida foram captados em ambiente doméstico (coleta passiva). Durante o período de coleta, os sintomas eram avaliados periodicamente através de questionários sMFQ, preenchidos a cada dois dias.

Nosso estudo investiga a depressão em adolescentes da base de dados IDEA-RiSCo. Exploramos quatro modalidades de dados coletados via *smartphones* como marcadores digitais de depressão: dados acústicos de relatos e episódicos, atividade e mobilidade. Esses dados foram agregados a cada dois dias e comparados aos escores da depressão usando o sMFQ.

Metodologicamente, a maioria dos estudos analisados buscam identificar comportamentos generalizáveis em um grupo de participantes (abordagem nomotética) e não consideraram o contexto e as particularidades de cada indivíduo (abordagem idiográfica), enquanto, optamos por um estudo comparativo entre modelos desenvolvidos em ambas as abordagens para investigar a depressão.

Além disso, propomos uma discussão sobre o potencial de uma série de marcadores digitais extraídos de *smartphones* e sua correlação com a sintomatologia da depressão na adolescência.

4 METODOLOGIA

Este Capítulo descreve a metodologia proposta na tese. São apresentados a organização dos conjuntos de dados, estratégias adotadas para extrair os atributos, processo de treinamento, medidas de avaliação e atividades desenvolvidas na tese.

4.1 Aspectos metodológicos

Esta tese propõe um modelo multimodal para identificar a depressão em adolescentes brasileiros. A Figura 6 apresenta a visão geral do modelo de AM multimodal de detecção de depressão estabelecido neste estudo. Ele contempla quatro modalidades de dados do IDEA-Tech: áudios de relatos, áudios episódicos, acelerômetro e GPS.

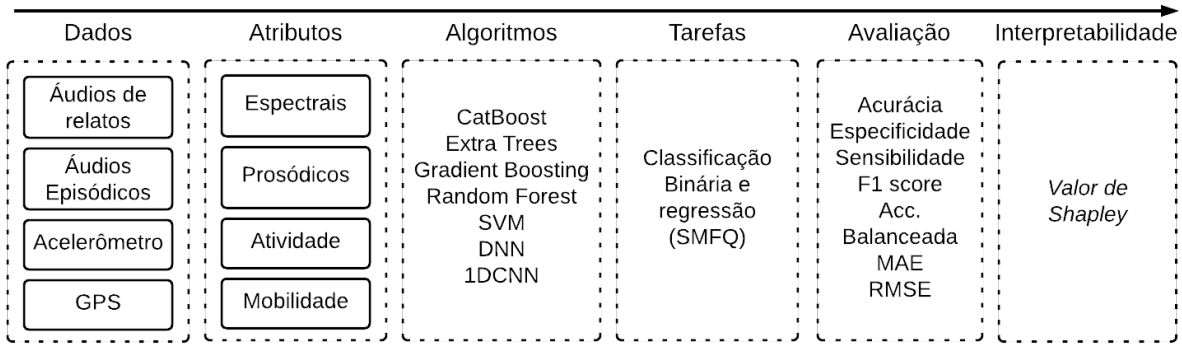


Figura 6 – Fluxograma do modelo multimodal de detecção de depressão. Fonte: Autoria própria

Nós adotamos estratégias de extração de atributos para cada modalidade. Além disso, avaliamos o desempenho de cinco algoritmos de aprendizado clássico e duas arquiteturas de aprendizado profundo, em tarefas de classificação binária e regressão, baseadas na pontuação do sMFQ preenchidos pelos adolescentes a cada dois dias. Utilizamos as medidas de Acurácia, Acurácia Balanceada, Especificidade, Sensibilidade, Medida-F1, MAE e RSME para avaliar o desempenho dos modelos. Ao final, medimos a importância dos atributos durante o treinamento, para compreender como eles impactam as previsões dos modelos preditivos.

Metodologicamente, nós adotamos duas formas de agregação de dados para avaliar potencial dos modelos preditivos de depressão: uma janela de agregação fixa, em que os dados foram agrupados a cada dois dias para calcular os atributos, e outra que acumula dados coletados em dias anteriores para calcular os atributos com base no sMFQ de referência.

Nós extraímos uma série de atributos acústicos, de atividade e mobilidade que foram integrados aos modelos preditivos, considerando a pontuação do sMFQ como verdade básica. Realizamos uma tarefa de classificação binária para identificar amostras deprimidas e não deprimidas dos adolescentes. Para tanto, nós consideramos que o adolescente está em estado deprimido caso tenha uma pontuação do sMFQ $\geq$ a 12. Além disso, a pontuação total do sMFQ é usada nos modelos de regressão para prever a intensidade dos sintomas da depressão.

Nós adotamos um método de importância para investigar quais os atributos ou combinação entre eles, possuem correlação mais forte com os sintomas da depressão. Os detalhes sobre as atividades de previsão do estado de humor atual e futuro serão apresentados no Capítulo 5.

4.1.1 Concepção dos conjuntos de dados

O IDEA-Tech é uma spin-off do IDEA-RiSCo, uma base de dados que contém 150 participantes adolescentes brasileiros com baixo risco, alto risco e diagnosticados com depressão que tiveram seus dados coletados via aplicativos instalados nos *smartphones Android* por um período de duas semanas.

Os conjuntos de dados adotados neste estudo incluem os áudios de relatos capturados por um chatbot (IDEA-Bot) utilizado para interação com os adolescentes através do *WhatsApp*; áudios episódicos, GPS e dados de acelerômetro coletados pelo aplicativo EBM instalado em *smartphones Android*, questionários sMFQ preenchidos pelos participantes, além da data e hora em que os dados foram coletados.

A Tabela 8 apresenta a quantidade de dados de atividade, localização e áudios episódicos coletados pelo EBM, dos áudios de relatos e questionários coletados pelo IDEA-Bot, durante o segundo ano (W2) do estudo IDEA-RiSCo.

Tabela 8 – Coleta de dados do IDEA-Tech (EBM e Bot) em números.

IDEA-Tech	Atividade	GPS	Episódicos	Relatos	MFQ	sMFQ
Amostras coletadas	2.450.406	2.839.011	1.039.354	5.434	-	-
Conjunto de dados (W2)	208.860	255.212	78.969	2.625	-	-
Adolescentes (W2)	104	105	102	112	110	668

As amostras que compõem o conjunto de dados do EBM no W2 foram coletadas

conforme a data em que o adolescente instalou o aplicativo no *smartphone* (D0), depois disso, os dados de cada adolescente foram coletados durante um período de 14 dias (D14). Em qualquer tempo os adolescentes puderam parar a coleta passiva de dados.

A média diária dos dados coletados pelo EBM por participante é de 166,58 amostras de GPS, 138,08 amostras de atividade e 54,25 áudios episódicos. No entanto, a quantidade de informações capturadas pelos aplicativos podem variar consideravelmente para cada adolescente.

Em alguns casos, o EBM deixou de coletar os dados de adolescentes por dias inteiros, especialmente na segunda metade do estudo. A diminuição do número de adolescentes com dados coletados ao longo dos dias foi possivelmente ocasionada pela desistência de alguns adolescentes do projeto. Além disso, um número reduzido de adolescentes possuem uma quantidade de dados coletados muito acima da média geral.

Nós utilizamos apenas os áudios coletados através da interação com o IDEA-Bot via *WhatsApp* e os dados coletados passivamente usando o aplicativo EBM. A Figura 7, apresenta o fluxo de coleta de dados desses aplicativos. Os dados utilizados nesta pesquisa foram coletados durante o W2.

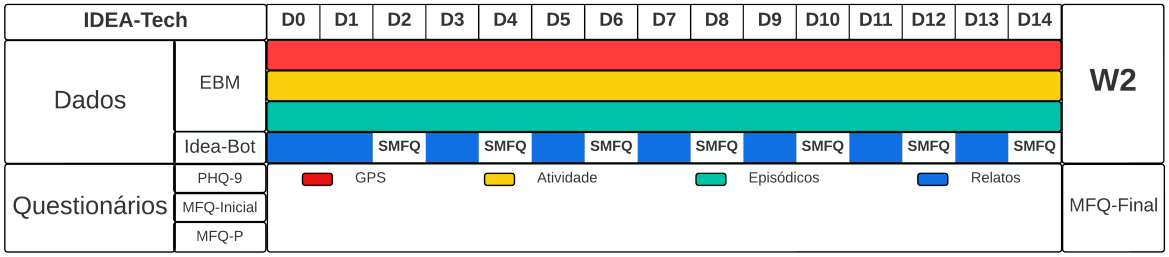
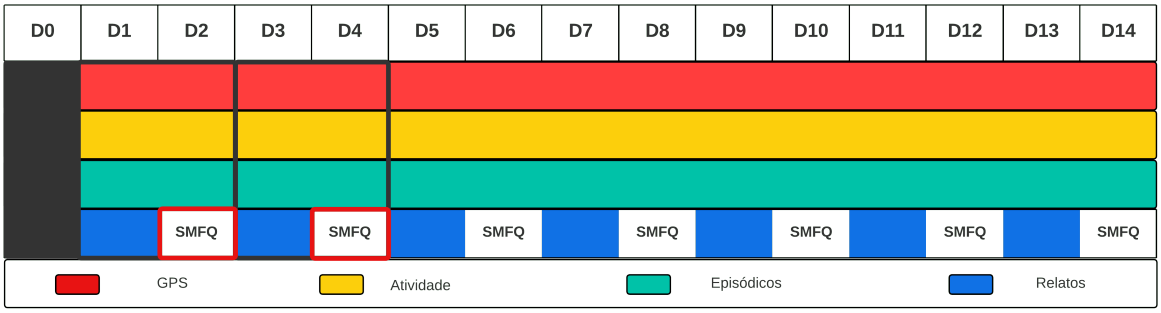


Figura 7 – Processo de coleta de dados do IDEA-Tech

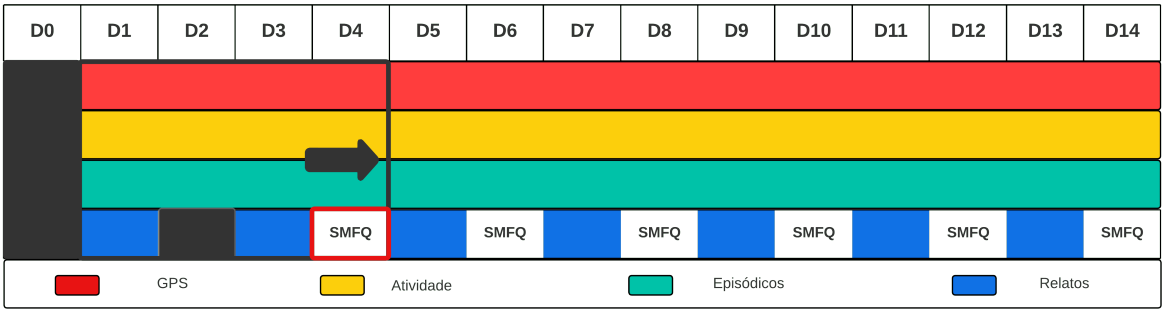
Antes do início do período de coleta, os adolescentes responderam aos questionários PHQ-9 e MFQ e seus pais ou responsáveis responderam ao MFQ-P. A coleta de dados começa em D0. O D0 é a data da interação inicial dos adolescentes com o aplicativo. Os áudios coletados pelo IDEA-Bot em D0, foram utilizados para testes da estrutura tecnológica e para a ambientação dos adolescentes.

Em seguida, o aplicativo EBM capturou passivamente os dados de GPS, atividade e áudios episódicos dos adolescentes por 14 dias. O IDEA-Bot coletou os áudios de relatos a cada dois dias. Os adolescentes responderam ao sMFQ em dias alternados à coleta dos áudios de relatos. No final do período de coleta, eles completaram um MFQ.

Alguns adolescentes ativaram a coleta de dados do EBM, em horários atípicos como, por exemplo, no meio da tarde do primeiro dia de coleta, captando um volume de dados reduzido que pode não refletir o comportamento diário dos adolescentes.



(a) Na agregação fixa, os dados são agrupados a cada dois dias para calcular os atributos. Exemplo: (D1+D2 => sMFQ1, D3+D4 => sMFQ2)



(b) Na agregação cumulativa os dados são agrupados sequencialmente usando os dias anteriores à data do sMFQ. Exemplo: (D1+D2 => sMFQ1, D1+D2+D3+D4 => sMFQ2)

Figura 8 – Agregações de dados fixa e cumulativa

Por essas razões, os dados coletados em D0 foram descartados da versão final do conjunto de dados.

4.1.2 Agregações de dados fixa e cumulativa

Os métodos de coleta passiva usados no IDEA-Tech resultaram em altas taxas de dados faltantes. No entanto, não implementamos métodos de imputação de dados neste estudo. Avaliamos duas formas de agregar os dados dos adolescentes ao longo dos dias (agregações de dados fixa e cumulativa) para calcular os atributos e mitigar o impacto dos dados ausentes.

Na abordagem de agregação fixa, agrupamos os dados dos adolescentes em uma janela de duração de dois dias com base na data de referência do sMFQ. Na abordagem de agregação cumulativa, agrupamos os dados dos adolescentes sequencialmente, considerando os dias anteriores à data de referência estabelecida pelo sMFQ.

As estratégias de agregação são apresentadas na Figura 8. É importante ressaltar que os questionários sMFQ não foram usados como variável de entrada dos modelos preditivos. Além disso, nenhuma informação coletada em dia posterior ao sMFQ de referência foi usada para calcular os atributos.

Os dados coletados durante o intervalo de duas semanas foram utilizados no trei-

namento dos algoritmos para avaliar o desempenho dos modelos na previsão do escore binário de depressão e da intensidade dos sintomas com base no questionário sMFQ. Os conjuntos de dados adotados neste estudo têm distribuições desequilibradas devido à menor quantidade de amostras da classe deprimida.

4.2 Estratégias adotadas para extrair os atributos

Nós extraímos uma série de atributos acústicos, de atividade e mobilidade coletados a partir dos *smartphones* dos adolescentes do IDEA-Tech. Nosso objetivo é explorar o potencial dessas informações como marcadores digitais da depressão.

Inicialmente, extraímos atributos espectrais, baseados no espectro de potência do sinal de fala, dos áudios de relatos capturados pelo IDEA-Bot. Esses atributos são obtidos convertendo o sinal de áudio do domínio do tempo para o domínio da frequência através da transformada de *fourier* e são bastante utilizados para classificar sons de uma forma geral.

Nós analisamos características associadas à presença e ausência de fala humana dos áudios episódicos. Para isso, extraímos atributos prosódicos dos áudios episódios para avaliar as interações sociais dos adolescentes em seu ambiente doméstico. Tais atributos são baseados na frequência fundamental (F0) e na duração dos sons vocais, não vocais e das pausas.

Nós também medimos a mobilidade dos adolescentes do IDEA-Tech usando cálculos de distâncias percorridas, entropias, agrupamentos, tempo de deslocamentos, entre outros.

Por fim, nós calculamos o tempo gasto pelos adolescentes em cada uma das atividades interpretadas pelo EBM e a frequência com que elas ocorreram para avaliar os níveis de atividade dos adolescentes. Todos os detalhes sobre os atributos utilizados adotados nesta tese podem ser consultados no Apêndice B.

Os conjuntos de dados adotados nesta tese são formados por 22 atributos espectrais; 18 atributos prosódicos; 21 atributos de mobilidade e 20 atributos de atividade. Assim, um total 81 atributos foram investigados como marcadores digitais de depressão na adolescência. As estratégias de pré-processamento e extração de atributos adotadas neste trabalho são discutidas adiante.

4.2.1 Atributos espectrais

Os áudios de relatos são enviados pelos adolescentes em resposta a um conjunto de perguntas relacionadas ao humor coletadas pelo IDEA-Bot através do *WhatsApp*. Eles foram pré-processados usando ferramentas de extração automática de atributos de áudio, consolidadas na literatura, *Librosa* (MCFEE et al., 2015) e Iracema (MAGALHAES; BARROS; LOUREIRO, 2020) para extrair os atributos espectrais dos áudios

de relatos.

Os Atributos espectrais são descritores de baixo nível (LLD - do inglês *Low Level Descriptors*) usados para representar o trato vocal e o som produzido na fala. Eles são obtidos transformando o sinal de fala do domínio do tempo para o domínio da frequência através da transformada de *Fourier* para isolar as frequências dominantes do sinal de áudio, facilitando sua interpretação por seres humanos ou sistemas computacionais.

Os espectrogramas e os Coeficientes Cepstrais de Frequência Mel (MFCC - do inglês *Mel-Frequency Cepstral Coefficients*) são exemplos de atributos espectrais utilizados com sucesso para investigar a fala deprimida conforme as análises conduzidas pelos estudos de Rejaibi et al. (2019); Dinkel et al. (2019); Alghowinem et al. (2020); Lin et al. (2020); Alghifari et al. (2019) e Lee et al. (2021).

O processo de extração dos atributos espectrais foi realizado com taxa de amostragem de 22.050, tamanho do quadro (n_fft = 2048), comprimento da janela FFT (win_length=n_fft), comprimento do salto (hop_length = 512) e transformada discreta de cosseno (DCT - do inglês *Discret Cosine Transform*) do tipo 2.

MFCC e espectrogramas foram amplamente investigados como marcadores digitais da fala deprimida. Contudo, combinamos uma série de atributos espectrais como taxas de cruzamento; energias e centróides espectrais e harmônicos; contrastes, fluxos, entropia, entre outros atributos baseados no espectro de potência do sinal que, até onde sabemos, não foram suficientemente explorados. A Tabela 9 apresenta o número de coeficientes utilizados para cada atributo espectral adotado neste estudo.

Tabela 9 – Número de coeficientes dos atributos espectrais

Coeficientes	Atributos
1	w2_bot_spe_zero_crossing_rate; w2_bot_spe_pitch; w2_bot_spe_rolloff; w2_bot_spe_entropy; w2_bot_spe_centroids; w2_bot_spe_spread; w2_bot_spe_skewness; w2_bot_spe_kurtosis; w2_bot_spe_energy; w2_bot_spe_rms_energy; w2_bot_spe_harmonic_energy; w2_bot_spe_harmonic_centroid; w2_bot_spe_noisiness; w2_bot_spe_hfc; w2_bot_spe_peak; w2_bot_spe_flux; w2_bot_spe_flatness
3	w2_bot_spe_bandwidth
6	w2_bot_spe_tonnals
7	w2_bot_spe_contrast
40	w2_bot_spe_mfcc_accelerate
128	w2_bot_spe_logMel_spectrograms

Os atributos acima são unidos sequencialmente em um vetor de comprimento

205 dimensões. Essas transformações são necessárias aos áudios de relatos para reduzir a complexidade e a dimensionalidade das características acústicas da fala (ALGHOWINEM et al., 2020). Ao mesmo tempo, também contribui para torná-los anônimos ou menos sensíveis à engenharia reversa.

4.2.2 Atributos prosódicos

Os áudios episódicos foram coletados pelo EBM para avaliar as interações sociais dos adolescentes em seu ambiente doméstico. Neste estudo, investigamos características de prosódia da fala dos áudios episódicos que estão relacionadas à presença e ausência de fala humana como marcadores digitais da depressão na adolescência.

Atributos prosódicos são as características da fala que podem ser percebidas pelos seres humanos. Eles representam os elementos da fala como sílabas, palavras e frases. As características prosódicas mais utilizadas em estudos de detecção de emoções na fala são baseados na energia, na duração da fala e na F0 (AKÇAY; OĞUZ, 2020; ALBUQUERQUE et al., 2021).

Nós extraímos características quantitativas da prosódia vocal contidas nos áudios episódicos. Especificamente, usamos a biblioteca *Disvoice* (VÁSQUEZ-CORREA et al., 2018), para extrair atributos prosódicos baseados na F0; e na duração dos sons vocais, não vocais e das pausas presentes nos áudios episódicos.

Em síntese, a F0 é um LLD capaz de emular a taxa de abertura e fechamento da glote, bastante utilizada para medir a vibração das cordas vocais durante a fala. Os sons vocais (*voiced*) são produzidos a partir da vibração das cordas vocais, enquanto os sons não vocais (*unvoiced*) são produzidos sem a vibração das cordas vocais (SHARMA; RAJPOOT, 2013). Finalmente, a pausa é a ausência de sons (silêncio) contida em um sinal de áudio.

Avaliamos 18 atributos prosódicos como marcadores digitais de depressão na fala. Usamos quatro medidas estatísticas (média, desvio padrão, assimetria e curtose) calculadas com base em características do contorno (amplitudes e entonação) da F0, na duração dos sons vocais e não vocais e das pausas contidas nos áudios episódicos.

4.2.3 Atributos de atividade

O EBM pré-processa os dados do acelerômetro de 3 eixos dos *smartphones* dos adolescentes no domínio do tempo para estimar a atividade que os adolescentes estão realizando no seu cotidiano no momento da coleta dos dados.

Esse processo é realizado transformando os valores numéricos dos eixos x, y e z do acelerômetro no domínio do tempo para seis atividades: "*Still, Unknown, On_Foot, Tilting, In_Vehicle e On_Bicycle*". Além disso, cada atividade tem um timestamp: 2020-08-27 13:42:08.678000+00:00" e um valor de confiança que pode variar de 0 a 100, representando a confiança do EBM de que determinada atividade detectada está re-

almente acontecendo no momento.

Os dados de atividade são em geral mais simples de pré-processar do que os dados descritos anteriormente. Basicamente, nós avaliamos 20 atributos de atividade como marcadores digitais de depressão na adolescência. Eles são baseados no tempo gasto pelos adolescentes em cada atividade e na frequência com que essas atividades ocorreram (foram capturadas pelo EBM).

4.2.4 Atributos de mobilidade

A mobilidade dos adolescentes é interpretada mediante um conjunto de coordenadas de latitude e longitude capturados temporalmente. Utilizamos bibliotecas especialistas como scikit-mobility (SKMOB) desenvolvida por Pappalardo et al. (2019), e GPS2Space (ZHOU et al., 2021), além de cálculos estatísticos independentes para extrair características espaço-temporais dos dados de localização.

Nós também extraímos características de mobilidade dos adolescentes, baseadas no agrupamento dos pontos de localização usando o algoritmo de agrupamento de dados espaciais *Density-Based Spatial Clustering of Applications with Noise* DBSCAN (ESTER et al., 1996), incorporado ao SKMOB.

Nós investigamos se características espaciais e temporais dos dados de localização podem estabelecer uma boa correlação com os sintomas da depressão na adolescência. Para tanto, estabelecemos quatro categorias de marcadores digitais de mobilidade baseados na distância, no tempo/deslocamento, no agrupamento dos locais e na área.

Atributos de mobilidade baseados na distância calculam a distância percorrida (em quilômetros) pelos adolescentes entre os pontos de localização. São calculados:

- Distância máxima percorrida
- Distância máxima percorrida em linha reta
- Distância máxima percorrida de casa ou local de referência

Nós estabelecemos como o local de referência o ponto de localização onde o adolescente permanece por mais tempo entre as 22:00h e 07:00h.

Atributos de mobilidade baseados no tempo/deslocamento medem o tempo médio de espera (em segundos) para que o adolescente se desloque entre os pontos de localização, bem como a distância média (em quilômetros) entre eles. Além disso, nós adotamos cálculos diversos de entropia e a variação de localização, que consideram a frequência, o tempo gasto e a ordem na qual os locais foram visitados.

- Tempo médio entre os deslocamentos
- Distância média entre os deslocamentos

A entropia é um princípio da termodinâmica que estima o grau de previsibilidade ou aleatoriedade de um sistema. Atributos baseados em entropia foram utilizados no contexto da depressão para avaliar a previsibilidade do *status* da tela, da conectividade com a internet e uso dos aplicativos dos *smartphones* (ASARE et al., 2021), do deslocamento espacial (SAEB et al., 2015, 2016) e (WANG; YALCIN; VANDEWEERD, 2020), do tempo gasto nos clusters de localização (MASUD et al., 2020), dos batimentos cardíacos (BYUN et al., 2019), da fala de pacientes deprimidos (YINGTHAWORNSUK, 2016), entre outras aplicações.

Nós avaliamos três variações da entropia como marcadores digitais de depressão associados a mobilidade dos adolescentes, são elas:

- Entropia aleatória
- Entropia não correlacionada
- Entropia real
- Variação da localização

A entropia aleatória mede o grau de previsibilidade do movimento dos adolescentes considerando que cada local visitado anteriormente tem a mesma probabilidade de ser selecionado como próximo local a ser visitado (SONG et al., 2010; PAUL et al., 2017). A entropia não correlacionada considera uma distribuição de probabilidades baseada na frequência na qual os locais foram visitados, mas ignora a relação temporal entre eles (SONG et al., 2010; WANG; YALCIN; VANDEWEERD, 2020). A entropia real considera a ordem espaço-temporal dos padrões de mobilidade dos adolescentes. Ela é calculada com base na frequência de visitação, na ordem na qual os locais foram visitados e o tempo gasto em cada local (PAPPALARDO et al., 2019; WANG; YALCIN; VANDEWEERD, 2020).

Adicionalmente, a variação de localização mede a variabilidade das coordenadas do GPS dos adolescentes. Ela é calculada como o logaritmo da soma das variações das coordenadas de latitude e longitude dos pontos de localização de um indivíduo (SAEB et al., 2015).

Os atributos de mobilidade baseados em agrupamentos avaliam a relação entre os pontos de localização com base na proximidade espacial entre eles. Estabelecemos um raio de 0,2 km na função "spatial_radius_km" do SKMOB e o número mínimo de 1 ponto de localização vizinho para formar os agrupamentos (*clusters*). Isso significa dizer que, os pontos de localização que estejam em uma distância de até 200 metros são considerados vizinhos e fazem parte do mesmo agrupamento. Com base nisso, nós quantificamos:

- O número total de visitas realizadas pelo adolescente;

- O número de locais distintos visitados pelos adolescentes
- O número de pontos de localização agrupados

O EBM realiza a coleta das coordenadas do GPS dos participantes de maneira aleatória e indiscriminada. A estrutura dos dados coletados pelo EBM não permite determinar de maneira natural, por exemplo, eventos como o início e o fim de uma caminhada, corrida ou viagem de carro. Por esse motivo estabelecemos caminhadas realizadas pelos adolescentes com base no agrupamento dos seus pontos de localização e nos locais de parada.

Nós determinamos um local de parada quando um adolescente permanece ao menos 20 minutos em um conjunto de locais situados a uma distância aproximada de 100 metros (0,5 x spatial_radius_km do SKMOB) durante o seu percurso diário. Os pontos de localização em cada local de parada são compactados em um único ponto, que contém as coordenadas medianas dos pontos agrupados e o tempo do ponto inicial.

Assim, uma caminhada é estabelecida como o percurso realizado pelos adolescentes entre os locais de parada durante o período de coleta. Com base nisso, nós estimamos:

- Total de caminhadas realizadas
- Número de caminhadas realizadas durante o dia
- Número de caminhadas realizadas a noite
- Média de caminhadas diária

Os atributos de mobilidade baseados na área medem a extensão geográfica por onde os adolescentes se deslocam no curso de suas atividades cotidianas em um período determinado de tempo.

Existem vários métodos espaciais e temporais para definir um espaço de atividade. Em geral, algumas características de dentro e fora dos espaços de atividades, como formato e tamanho, podem refletir comportamentos espaciais, sociais, de saúde física e mental das pessoas (SMITH; FOLEY; PANTER, 2019).

Nós utilizamos dois métodos para calcular os espaços de atividades (em metros quadrados) dos adolescentes:

- Espaços de atividade baseado no método de casco convexo.
- Espaços de atividade baseado no método de buffer.

O método de casco convexo, calcula os espaços de atividade com base na área do polígono formado pelos pontos de localização mais extremos do mapa de coordenadas do adolescente. O método de buffer, mede o espaço de atividade com base na área da circunferência dos locais visitados pelo adolescente, para isso foi estabelecido a distância do buffer (raio de 100 metros) para agrupar e dissolver os pontos de localização.

Além disso, nós adotamos duas medidas de raio de giro, que calculam a distância percorrida pelos adolescentes, induzida pelos locais visitados com maior frequência. São elas:

- Raio de giro
- Raio de k giro (onde k é igual aos 5 locais visitados com maior frequência pelo adolescente.)

O raio de giro mede a distância que um indivíduo se move em torno do seu centro da massa, sendo calculado a partir de um ponto médio ponderado de seus pontos de localização. Desse modo, indivíduos que possuem um raio de giro menor frequentam locais mais próximos uns dos outros que indivíduos que possuem um raio de giro maior (PAPPALARDO et al., 2015, 2016).

Os atributos de mobilidade estabelecidos nesta tese foram extraídos a partir de amostras de dados de localização estacionários, conforme descrito nos estudos de Saeb et al. (2015, 2016). Por esse motivo, nós adotamos um filtro para descartar os dados cuja velocidade estimada excedeu o limite de 1 km/h.

4.3 Avaliação do Modelo proposto

Nós adotamos uma tarefa de classificação binária para discriminar amostras deprimidas e não deprimidas com base na nota de corte do sMFQ e uma tarefa de regressão para identificar a pontuação exata do questionário sMFQ respondido pelos adolescentes.

Nós consideramos as métricas de Acurácia, Especificidade, Sensibilidade, Medida-F1 (F1), Acurácia Balanceada, MAE e RMSE para avaliar o desempenho dos modelos de detecção de depressão.

Estabelecemos os melhores modelos preditivos durante os experimentos considerando as medidas de Acurácia Balanceada e MAE (quando houver), nessa ordem de importância. Além disso, a matriz de confusão foi gerada para os valores das classes positivas e negativas.

4.3.1 Sobreamostragem da classe minoritária

Nós utilizamos métodos de sobreamostragem da classe minoritária no conjunto de dados de treinamento, validação cruzada e otimização de hiperparâmetros para melhorar o desempenho dos modelos preditivos e torná-los menos suscetíveis à *overfitting*. Adotamos uma solução mista de sobreamostragem usando o *Borderline* SMOTE do tipo 1 e *RandomOverSampler*.

Basicamente, o algoritmo *Borderline* SMOTE1 calcula a distância de cada amostra da classe minoritária para seus vizinhos. Somente as amostras que possuem vizinhos da classe majoritária (*borderlines*) são reamostradas sinteticamente a partir de um processo de interpolação até igualar a quantidade de exemplos entre as classes (HAN; WANG; MAO, 2005).

O problema é que o *Boderline* SMOTE1 não consegue sobreamostrar algumas amostras de classes minoritárias, sobretudo na tarefa de regressão, devido à baixa quantidade de amostras disponíveis de algumas pontuações do sMFQ, em nosso conjunto de dados. Para esses casos, utilizamos o *RandomOverSampler* para replicar as amostras de classes minoritárias até atingir o equilíbrio entre todas as classes.

4.3.2 Importância dos atributos

Nós adotamos um método de importância baseado no algoritmo SHAP para identificar a contribuição dos atributos para prever os escores binários do sMFQ dos participantes do IDEA-Tech.

Os resultados são apresentados em forma de gráficos para destacar os 20 principais atributos que otimizaram os modelos preditivos que alcançaram o melhor desempenho nos experimentos.

O método de importância dos atributos favoreceu a interpretação dos modelos preditivos desenvolvidos nesta tese. Adotamos dois gráficos SHAP. Eles foram aplicados nos conjuntos de testes para quantificar a importância global (gráfico de barras) dos atributos, medida através dos valores médios de suas contribuições para prever os escores da depressão. Nós, também, medimos a direcionalidade dos atributos por meio de dois gráficos de resumo, identificando o impacto de cada atributo para a previsão das classes na saída do modelo preditivo.

As pontuações de importância forneceram informações relevantes sobre quais atributos mais influenciaram a predição dos escores do sMFQ. Eles permitiram melhorar o desempenho dos modelos preditivos durante o treinamento, sendo fundamentais para compreendermos a relação dos atributos com a sintomatologia da depressão na adolescência.

4.4 Atividades propostas na tese para investigar a depressão na adolescência

Neste estudo propomos duas atividades para investigar a depressão na adolescência que envolvem a previsão do humor deprimido atual e futuro dos adolescentes do IDEA-RiSCo.

Na primeira, realizamos um experimento de caráter exploratório para avaliar algoritmos de AM usando diferentes agregações de dados e combinações de modalidades para prever o estado de humor atual dos adolescentes com base no questionário sMFQ. Na segunda, desenvolvemos dois experimentos que avaliam a capacidade dos modelos usando as mesmas agregações e duas abordagens metodológicas para prever o humor deprimido em amostras futuras de questionários sMFQ respondidos pelos adolescentes. Ambas são detalhadas nas seções adiante:

4.4.1 Previsão do humor deprimido atual

Esta atividade tem o objetivo de comparar o desempenho de algoritmos de AM para prever o estado de humor atual dos adolescentes no momento da resposta ao questionário sMFQ que foi preenchido a cada dois dias.

Os adolescentes do IDEA-Tech foram orientados a responder o questionário sMFQ sobre o seu estado de humor atual. Especificamente, os modelos foram avaliados pela capacidade de prever o estado de humor dos adolescentes com base em dados coletados no dia anterior e no dia da resposta do questionário sMFQ.

Nós conduzimos os experimentos em duas configurações: unimodal e multimodal. Nos experimentos unimodais, os modelos de detecção de depressão são treinados para cada modalidade, individualmente. Foram realizados quatro experimentos unimodais, um experimento usando apenas atributos espectrais; utilizando apenas atributos prosódicos; usando apenas atributos de atividade; e por fim, um quarto, usando apenas os dados de mobilidade.

Os experimentos multimodais são formados a partir das combinações entre as modalidades utilizadas na etapa anterior. Assim, nós realizamos um total de 30 experimentos para avaliar o estado de humor atual dos adolescentes, quatro experimentos unimodais e onze experimentos multimodais para cada agregação de dados.

4.4.1.1 Design do modelo proposto

A Figura 9 traz uma visão geral do modelo proposto. Calculamos os atributos espectrais, prosódicos, de atividade e mobilidade conforme a data da resposta do questionário sMFQ e o tipo de agregação de dados. Os experimentos foram realizados para avaliar a eficácia dos modelos em prever a ausência ou presença de depressão com base na nota de corte binária do sMFQ.

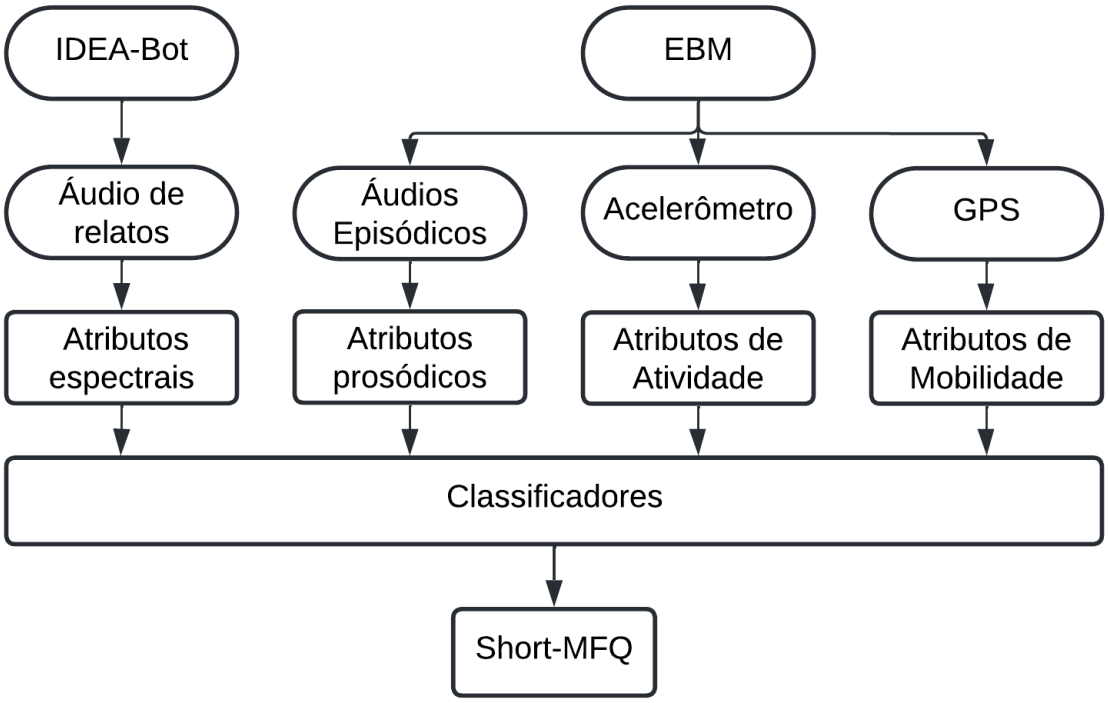


Figura 9 – Organização dos experimentos da atividade de previsão do humor deprimido atual: Fonte: Autoria própria

Os experimentos incluíram algoritmos de AM clássico e profundo, passaram pelas mesmas etapas de treinamento e avaliação, sob os mesmos hiperparâmetros de configuração.

Além disso, usamos um método de importância para identificar quais os atributos que melhor contribuíram para otimizar o desempenho do modelo preditivo de melhor resultado entre os experimentos unimodais e multimodais.

Desse modo, pudemos explorar nesta tese, além de métricas de desempenho, os atributos que mais contribuíram para otimizar o desempenho do modelo na tarefa de classificação da depressão.

4.4.1.2 Divisão dos conjuntos de dados de treinamento e teste

Existe uma grande variabilidade na coleta de dados do IDEA-Tech. Isso significa dizer que a frequência de coleta dos dados diária varia bastante entre os adolescentes, sobretudo devido à coleta passiva do EBM.

No entanto, para esta atividade, não há restrições relacionadas à quantidade de dados coletados ou sMFQ respondidos. Incluímos adolescentes que tiveram uma amostra válida (dois dias de dados coletados e sMFQ correspondente ao período de coleta) até adolescentes que realizaram o período completo da coleta e preencheram sete sMFQ.

A Figura 10 apresenta a distribuição dos adolescentes e amostras válidas, segundo a nota de corte binária para depressão do questionário sMFQ. Incluímos 514 amostras de 92 adolescentes e 582 amostras de 93 adolescentes para compor nossos conjuntos de dados agregados. Ambos, são severamente desequilibrados, possuindo cerca de 15% de amostras deprimidas. Além disso, a maioria das amostras deprimidas são do sexo feminino.

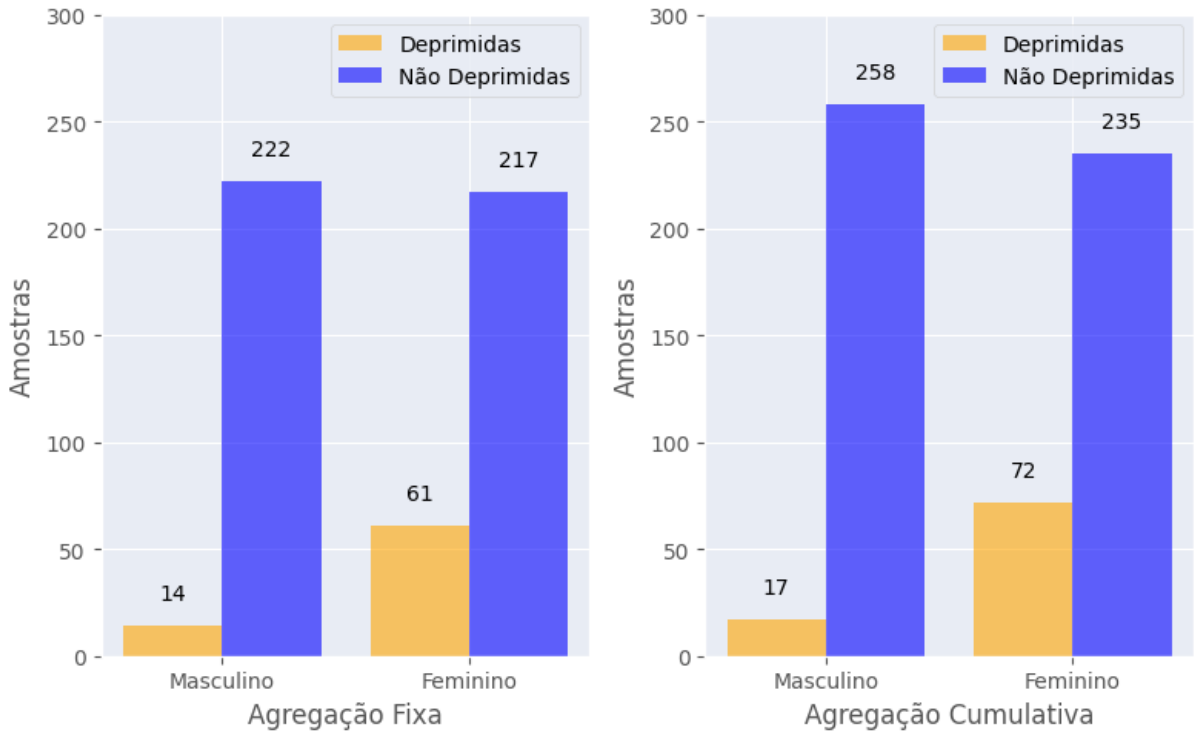


Figura 10 – Quantitativo dos conjuntos de dados adotados na atividade de previsão do humor deprimido atual.

As amostras foram divididas aleatoriamente em uma proporção de 75% para o conjunto de treinamento e 25% para o conjunto de testes, com divisão estratificada das amostras para garantir uma distribuição equilibrada de classes em ambos os conjuntos de dados. Dessa forma, os dados foram agrupados e pré-processados para treinar os classificadores com base nos dados coletados a cada dois dias e compará-los aos escores binários da depressão do sMFQ. O mesmo conjunto de testes foi usado para avaliar os modelos preditivos.

4.4.1.3 Busca aleatória com validação cruzada estratificada

Nós adotamos um método de busca aleatória de hiperparâmetros em 10 iterações usando o *RandomizedSearchCV*, com validação cruzada estratificada de 5 vezes para generalizar o processo de treinamento e reduzir problemas de superajuste dos modelos. A Figura 11 detalha o processo de treinamento dos modelos preditivos para esta atividade.

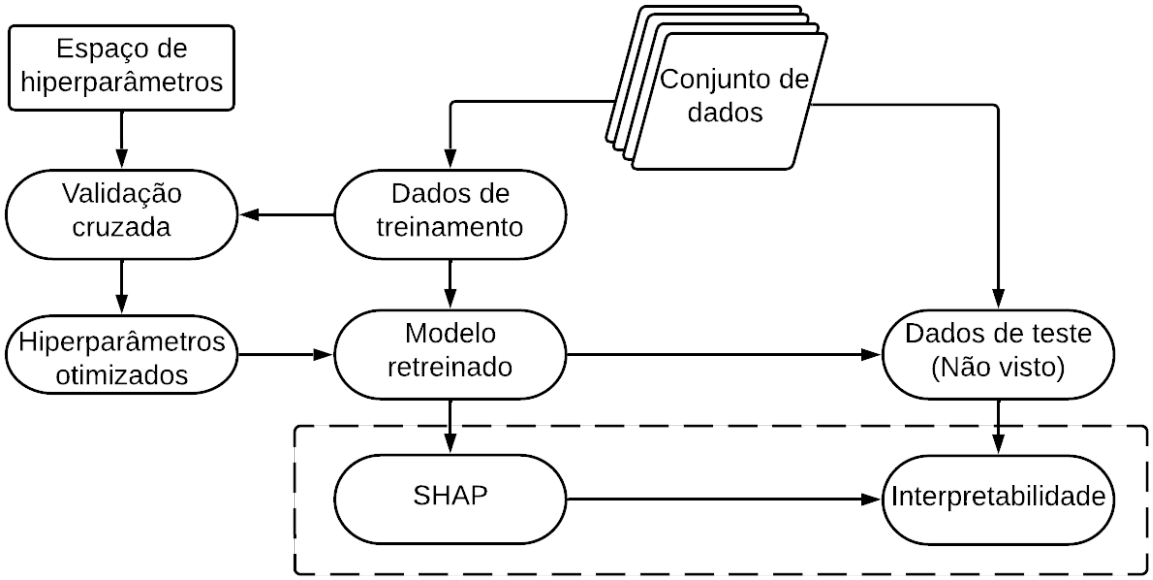


Figura 11 – Fluxograma do treinamento dos modelos de detecção de depressão. Dividimos os dados em 75% para treinamento e 25% para teste. Dez conjuntos de hiperparâmetros são selecionados aleatoriamente e avaliados durante a validação cruzada de 5 vezes. O modelo com o melhor desempenho, dado pela Acurácia Balanceada média no conjunto de validação, é retreinado em todo o conjunto de treinamento e avaliado uma única vez no conjunto de testes. Paralelamente, o algoritmo de *Shapley* computa os valores das contribuições dos atributos para as previsões dos modelos no conjunto de testes. Fonte: Autoria própria

Estabelecemos o número de dez interações (n_iter = 10) considerando os recursos computacionais e tempo gasto durante o treinamento para controlar o número de combinações de hiperparâmetros avaliados durante a busca aleatória (PATEL; KHALAF; AIZENSTEIN, 2016; GAO; CALHOUN; SUI, 2018).

O método de validação cruzada estratificada de 5 vezes (k=5) divide o conjunto de dados de treinamento em cinco partes de mesmo tamanho, em que quatro partes são combinadas para formar o conjunto de dados de treinamento, a outra parte é utilizada para a validação. Esse processo é executado para os modelos serem avaliados em um conjunto de dados de validação diferente a cada iteração (k) e cada modelo foi ajustado 50 vezes durante o processo de treinamento.

Nós usamos a Acurácia Balanceada média (B. Acc média) sobre o conjunto de dados de validação para determinar os valores ótimos dos hiperparâmetros dos modelos desenvolvidos. Ao final, o modelo com o melhor desempenho é retreinado com todo o conjunto de treinamento e avaliado uma única vez com conjunto de testes (não visto).

4.4.2 Previsão do humor deprimido futuro.

Este experimento tem por objetivo comparar diferentes algoritmos de AM clássico e profundo na previsão do humor deprimido em amostras futuras coletadas pelos adolescentes. Para isso, propomos duas atividades de detecção de depressão conside-

rando a ordem temporal na qual os aplicativos coletaram os dados dos adolescentes.

Na primeira atividade, dividimos os dados dos adolescentes semanalmente para avaliar a eficácia dos modelos preditivos. Esses modelos foram treinados com os dados coletados na primeira semana para prever as amostras obtidas na segunda semana do projeto. A segunda atividade avalia a capacidade dos modelos, treinados com amostras atuais, de prever as respostas futuras dos adolescentes aos sMFQ.

Além disso, comparamos modelos de AM multimodais baseados em duas abordagens de pesquisa científica, (nomotética e idiográfica), amplamente utilizadas nas ciências sociais em pesquisas que visam compreender o comportamento humano.

A abordagem nomotética investiga padrões comuns de comportamento humano entre um determinado grupo de pessoas. Ela identifica relações causais e comportamentos generalizáveis que contribuem para a depressão e podem ser explorados em outras populações ou grupos de indivíduos deprimidos (SALVATORE; VALSINER, 2010; BELTZ et al., 2016).

A abordagem idiográfica se concentra na complexidade e nas especificidades do comportamento humano. Ela permite uma compreensão mais aprofundada e contextualizada da experiência do indivíduo, seus sentimentos, pensamentos, entre outros aspectos investigados na depressão (BELTZ et al., 2016; AALBERS et al., 2023).

É importante enfatizar que ambas as abordagens têm vantagens e limitação, mas também podem ser utilizadas de maneira complementar, aprimorando a compreensão sobre comportamento humano associado a depressão (JACOBSON; CHUNG, 2020). A seguir, apresentamos os modelos nomotéticos e idiográficos de detecção de depressão desenvolvidos para este estudo.

4.4.2.1 Design do modelo proposto

Nós introduzimos um novo modelo preditivo de depressão, multitarefa e multimodal, baseado em uma rede neural convolucional de 1 dimensão (MT1DCNN) que contempla simultaneamente as tarefas de classificação binária e regressão (Figura 12).

Nós comparamos os algoritmos de AM clássicos (RF, ET, GB, CB, SVM) e profundos (DNN, 1DCNN e MT1DCNN). Avaliamos o desempenho dos algoritmos em duas tarefas de detecção de depressão: uma tarefa para separar as amostras deprimidas e não deprimidas dos adolescentes, e outra tarefa para estimar a intensidade dos sintomas de depressão. Ambas foram baseadas nas pontuações do questionário sMFQ respondidos pelos adolescentes.

Os experimentos nomotéticos e idiográficos combinam as modalidades do experimento multimodal ARM (Atividade, Áudios de relatos e Mobilidade). Além disso, realizamos experimentos usando agregação fixa e cumulativa conforme atividade anterior.

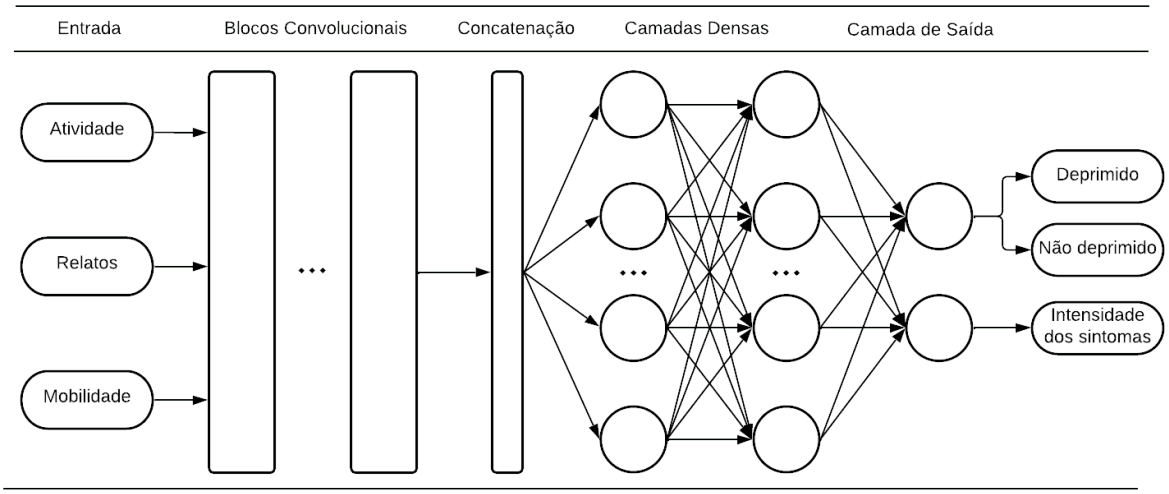


Figura 12 – Modelo multitarefa MT1DCNN: Os hiperparâmetros incluem o número de camadas convolucionais e densas, o número de neurônios, taxas de *dropout* e aprendizado, entre outros.

4.4.2.2 Divisão dos conjuntos de dados de treinamento e teste

As Figuras 13 e 14 apresentam a distribuição das amostras válidas dos adolescentes usadas em nossos conjuntos de dados conforme o ponto de corte binário e a pontuação total do sMFQ, respectivamente. Utilizamos os mesmos conjuntos de dados para as duas atividades de previsão do humor deprimido futuro.

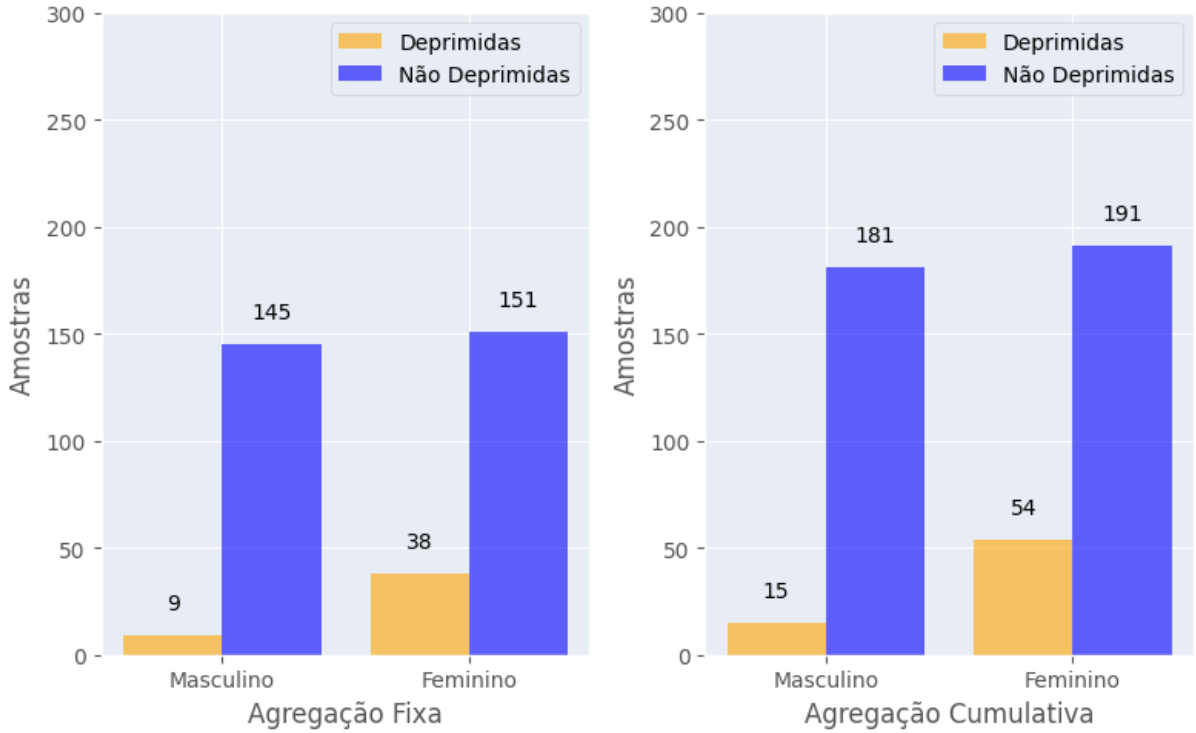
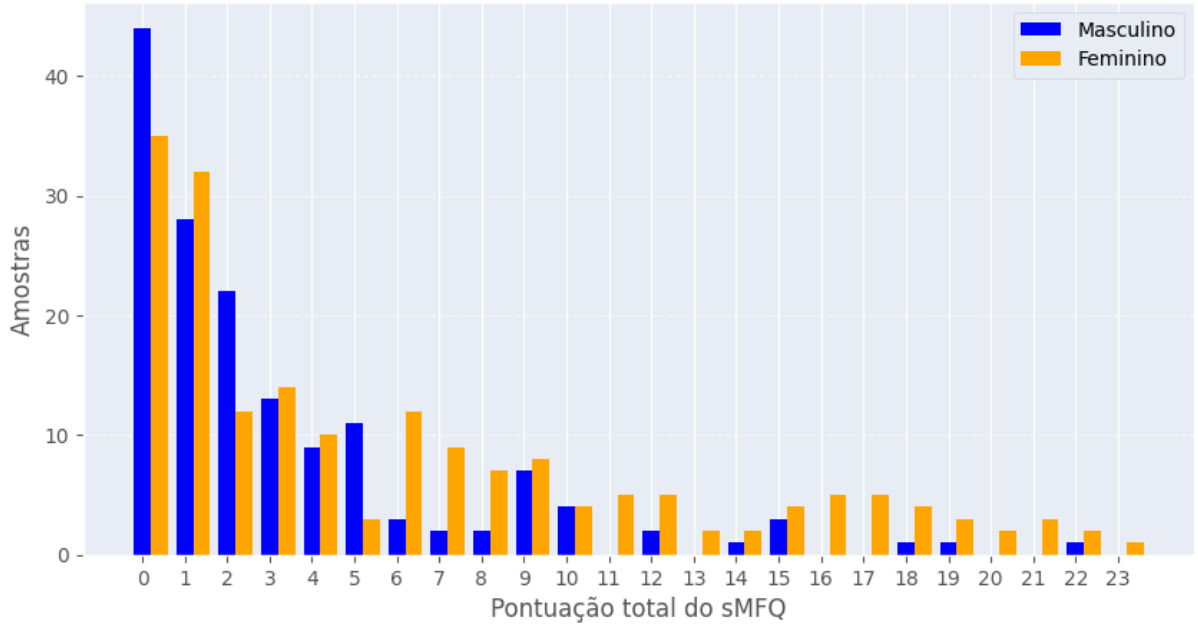
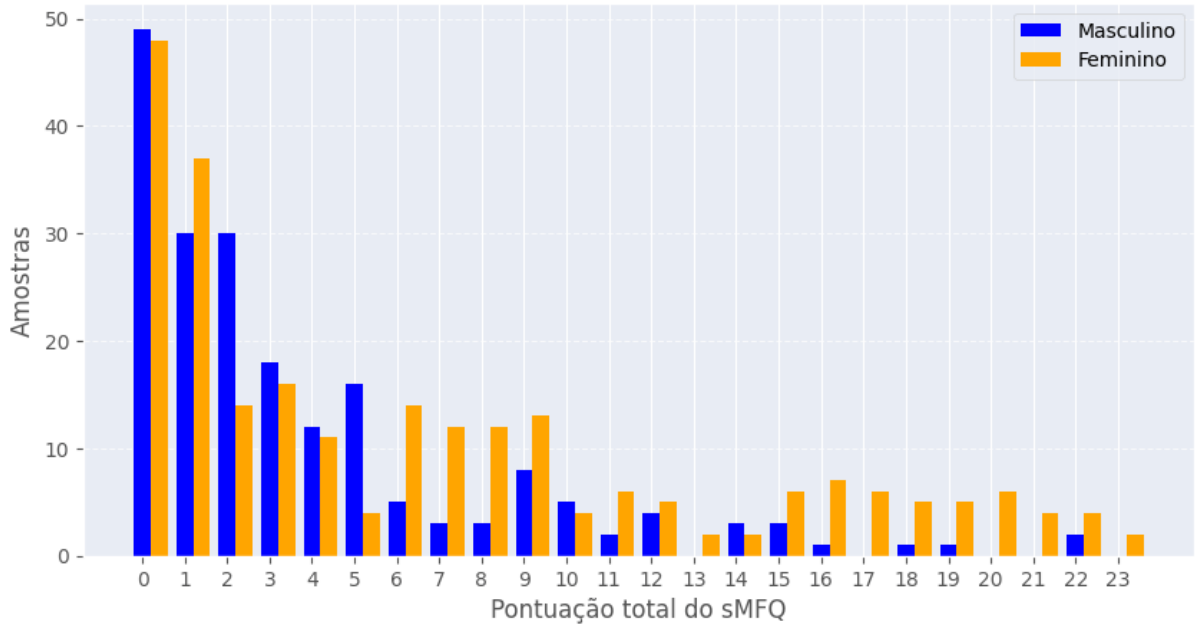


Figura 13 – Quantitativo dos conjuntos de dados na tarefa de classificação binária

Nós estabelecemos um critério de inclusão para controlar a quantidade de amos-



(a) Agregação fixa



(b) Agregação cumulativa

Figura 14 – Quantitativo dos conjuntos de dados na tarefa de regressão

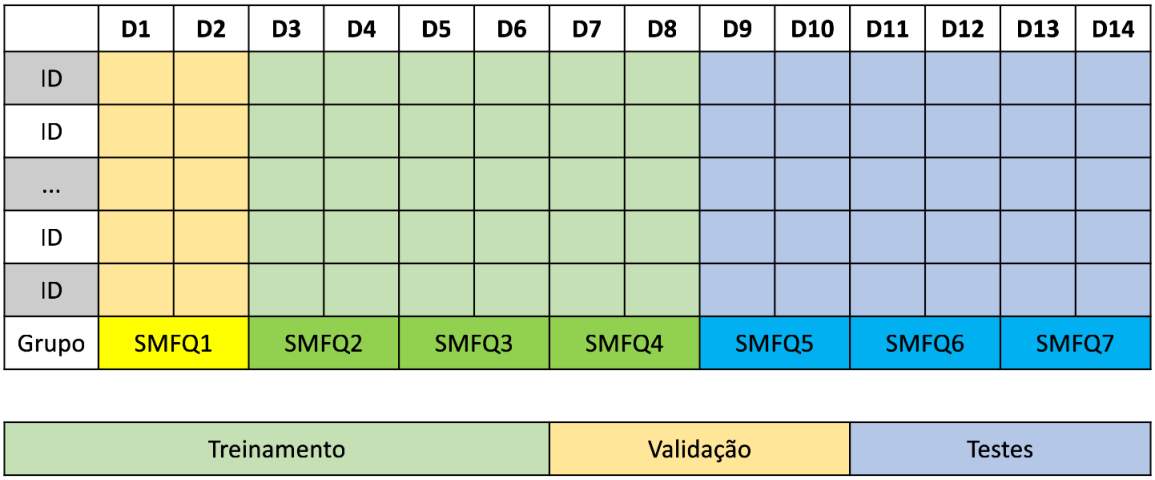
tras e adolescentes envolvidos na atividade. Portanto, apenas os adolescentes que completaram o período total da coleta (14 dias) e responderam os sete sMFQ formaram os conjuntos de dados utilizados nos experimentos para prever estados deprimidos futuros. Além disso, as amostras foram divididas sequencialmente, seguindo a ordem cronológica na qual os aplicativos coletaram os dados.

Nossos conjuntos de dados são formados 343 amostras de 49 adolescentes usando agregação fixa e 441 amostras de 63 adolescentes usando agregação cu-

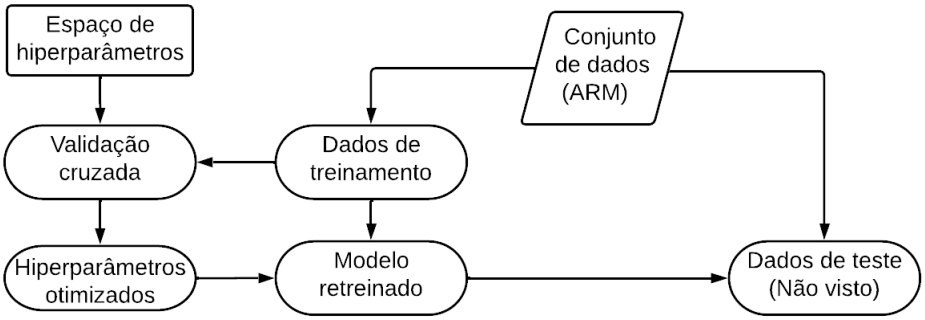
mulativa. Ambos, também se mostraram bastante desequilibrados, possuindo cerca de 15% de amostras deprimidas, em sua maioria do sexo feminino.

4.4.2.3 Previsão de amostras da próxima semana

Nós aplicamos uma estratégia de validação cruzada LOGO para agrupar as amostras com base na ordem na qual os sMFQ foram respondidos pelos adolescentes. A Figura 15a detalha o processo de validação cruzada de grupos adotada nesta atividade. Metodologicamente, os modelos nomotéticos utilizaram as mesmas etapas de treinamento do experimento anterior (Figura 15b).



(a) Validação cruzada deixa um grupo de sMFQ de fora



(b) Fluxograma de treinamento dos experimentos nomotéticos

Figura 15 – Treinamento dos modelos nomotéticos. A cada interação da validação cruzada, um grupo de sMFQ é selecionado para validação enquanto os outros grupos de sMFQ são combinados no conjunto de treinamento do modelo preditivo. Esse processo se repete até que todos os grupos sMFQ sejam selecionados uma vez para o conjunto de validação. O modelo com melhor desempenho é determinado pela média da Acurácia Balanceada no conjunto de validação para a classificação binária e pela média do MSE no conjunto de validação para a regressão. Ao final, o modelo selecionado é retreinado em todo o conjunto de treinamento com amostras coletadas durante a primeira semana e avaliado uma única vez no conjunto de testes que usa dados coletados na segunda semana do período de coleta.

Os modelos treinados individualmente para cada tarefa usaram o método de busca

aleatória de hiperparâmetros de 10 repetições usando o *RandomizedSearchCV* seguindo a mesma metodologia descrita anteriormente. Usamos a biblioteca Optuna para realizar a otimização de hiperparâmetros do modelo multitarefa MT1DCNN.

Em síntese, o modelo MT1DCNN usa um método de validação cruzada para deixar um grupo de sMFQ de fora com otimização Bayesiana de hiperparâmetros. Selecionamos o modelo com melhor desempenho com base nos valores médios de Acurácia Balanceada e MSE durante a validação cruzada. Então, a função *objective* do Optuna retorna a diferença negativa entre as duas medidas, buscando durante a validação cruzada o modelo que melhor maximize a Acurácia Balanceada e minimize o MSE. Esse processo é realizado em 20 interações utilizando o algoritmo de otimização Estimador de Parzen estruturado em árvore (TPE - do inglês *Tree-structured Parzen Estimator*) (AKIBA et al., 2019; BERGSTRA et al., 2011). Ao final, o modelo é retreinado em todo o conjunto de dados de treinamento (amostras coletadas na primeira semana) e avaliado uma única vez no conjunto de dados de testes (amostras coletadas na segunda semana).

O experimento idiográfico segue o mesmo processo de treinamento do experimento nomotético. Porém, nós introduzimos três matrizes de similaridade para personalizar o modelo multitarefa MT1DCNN. A Figura 16 apresenta o fluxograma de treinamento dos experimentos idiográficos.

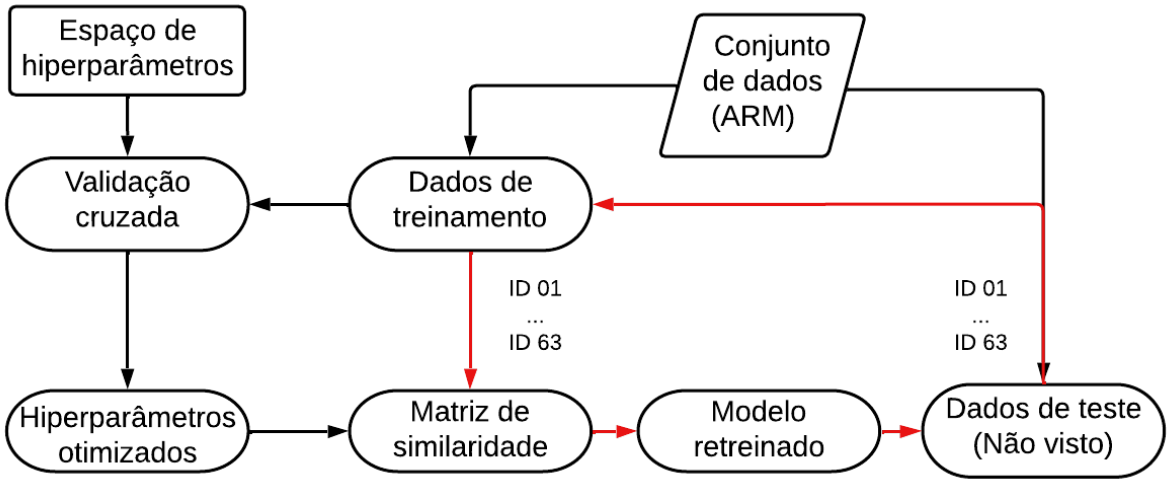


Figura 16 – Treinamento dos modelos idiográficos: O experimento idiográfico replica o mesmo processo de divisão de dados, validação cruzada e seleção de hiperparâmetros otimizados do experimento nomotético. A partir daí, treinamos novamente o modelo com uma matriz de similaridade responsável por atribuir pesos a todas as amostras do conjunto de treinamento com base em sua proximidade com as amostras do adolescente de referência (maiores pesos para amostras mais próximas). Em seguida, usamos o modelo, treinado com o vetor de peso personalizado, para prever as mesmas amostras de adolescentes no conjunto de teste. Esse processo se repete para cada adolescente.

Uma matriz de similaridade quantifica a semelhança ou proximidade entre as amostras de um conjunto de dados. Estudos anteriores adotaram matrizes de simila-

ridade em imagens de ressonância magnética para classificar Alzheimer (BEHESHTI et al., 2017) e autismo (LEMAITRE; NOGUEIRA; ARIDAS, 2017). Elas também foram usadas para classificar emoções em textos do twitter (BENROUBA; BOUDOUR, 2023).

Nosso estudo adotou matrizes para encontrar a similaridade entre os pontos de dados e atribuir pesos a todas as amostras dos conjuntos de dados, personalizando o treinamento do modelo idiográfico (MT1DCNN) com base nas amostras do adolescente de referência.

Criamos três matrizes de similaridade usando o método de similaridade de cosseno (MT1DCNN-C), distância euclidiana (MT1DCNN-E) e similaridade de Jaccard (MT1DCNN-J). Além disso, calculamos os valores absolutos das medidas de Acurácia, MAE e RMSE para cada adolescente e as medidas gerais para cada modelo idiográfico. Usamos medidas gerais para comparar o desempenho de modelos preditivos idiográficos e nomotéticos.

Nós adotamos o Optuna e otimização Bayesiana nos experimentos que envolveram o nosso modelo multitarefa MT1DCNN por que esse método de otimização consegue lidar com modelos de múltiplas entradas e múltiplas saídas de maneira mais eficiente que otimizadores tradicionais como *GridsearchCV* e *RandomizedSearchCV*, porém há um maior custo computacional envolvido.

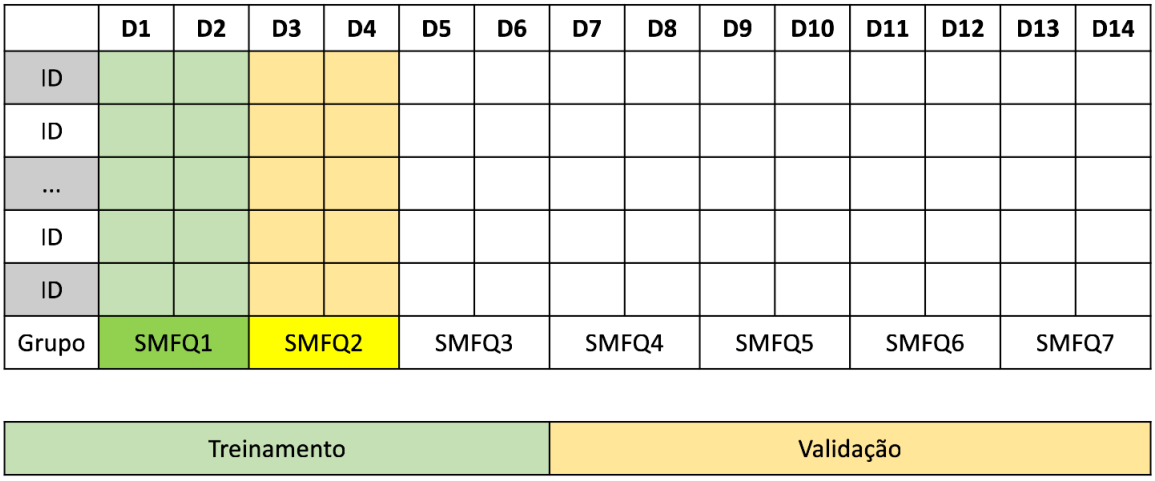
4.4.2.4 Previsão dos próximos sMFQ

Esta atividade foca na utilização dos questionários sMFQ respondidos durante o ciclo atual de coleta de dados para a previsão do humor deprimido futuro dos adolescentes.

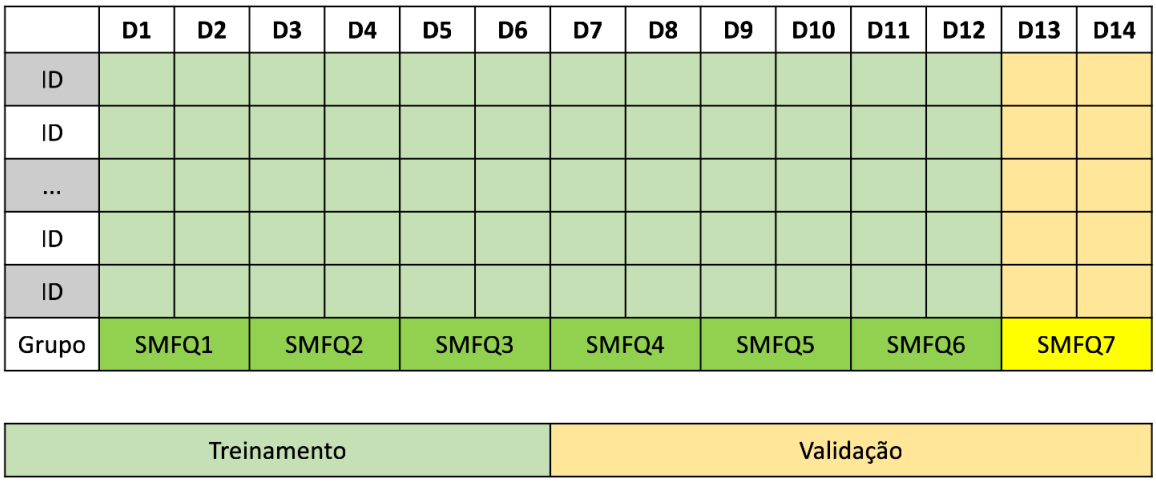
Durante o ciclo completo de coleta de dados, que corresponde a um período de 14 dias, os adolescentes responderam um total de 7 questionários SMFQ. O objetivo desta atividade é comparar modelos de previsão de depressão que nos permitam estimar o humor deprimido futuro dos adolescentes com base em seus padrões de respostas aos questionários sMFQ e dados históricos coletados pelos aplicativos.

Para tanto, dividimos as amostras sequencialmente utilizando um método de validação cruzada de séries temporais *TimeSeriesSplit* da biblioteca *Scikit-learn* (PEDREGOSA et al., 2011) para agrupar as amostras dos adolescentes com base na sequência em que foram coletadas.

Nós utilizamos os dados coletados nos dias anteriores ao sMFQ para treinar os modelos, respeitando a sequência dos dados em cada dobra de validação cruzada subsequente. Em seguida, usamos o mesmo método de otimização bayesiana de hiperparâmetros do Optuna, adotado nos experimentos multitarefas deste estudo. A Figura 17 detalha o processo de validação cruzada de séries temporais adotada nesta atividade.



(a) Previsão dos questionários sMFQ 2



(b) Previsão dos questionários sMFQ 7

Figura 17 – Validação cruzada de séries temporais

A cada interação da validação cruzada, nós escolhemos um grupo de sMFQ para validação, a partir do sMFQ2 (Figura 17a) enquanto os outros grupos de sMFQ anteriores ao sMFQ do conjunto de validação, são combinados no conjunto de treinamento do modelo preditivo. O processo se repete, sequencialmente, para garantir a ordenação natural dos dados de entrada no treinamento dos modelos preditivos, até que o último grupo (sMFQ7) seja selecionado para o conjunto de validação (Figura 17b). Em cada dobra, nós aplicamos uma função para sobreamostrar a classe minoritária, equilibrando as amostras do conjunto de treinamento. Determinamos o melhor modelo através da média da Acurácia Balanceada nos conjuntos de validação para a classificação binária e pela média do MSE nos conjuntos de validação para a regressão.

Nós comparamos os modelos preditivos nas tarefas de classificação binária e regressão usando a mesma combinação de modalidades (Experimento Multimodal ARM) e os mesmos conjuntos de dados agregados usados para a previsão de amos-

tras da próxima semana.

O treinamento dos experimentos nomotéticos e idiográficos também segue a mesma configuração da atividade anterior. No entanto, os experimentos de previsão dos próximos sMFQ são computacionalmente mais caros se comparado às atividades anteriores, por esse motivo, nós avaliamos o desempenho dos melhores modelos preditivos multitarefa, nomotético (MT1DCNN) e idiográfico (MT1DCNN-E), desenvolvidos na atividade anterior. Devido a essa limitação optamos por não calcular a importância dos atributos para esta atividade.

4.5 Experimentos

Nós avaliamos cinco algoritmos de aprendizado clássico: CB, ET, GB, RF e SVM e dois algoritmos de aprendizado profundo: DNN e 1DCNN. Mais detalhes sobre os algoritmos e seus conjuntos de hiperparâmetros podem ser consultados no Apêndice D.

Nós organizamos os resultados do estudo em duas etapas. A primeira etapa avalia o desempenho dos modelos preditivos para as tarefas de detecção de depressão propostas na tese. A segunda etapa avalia a qualidade dos atributos utilizados para prever a depressão na adolescência com base nos experimentos realizados na etapa anterior. Os modelos de detecção de depressão foram avaliados usando medidas de Acurácia (Acc), Especificidade (Esp), Sensibilidade (Sen), Medida-F1 (F1) e Acurácia Balanceada (B. Acc), MAE e RMSE.

Os modelos foram desenvolvidos usando a plataforma *TensorFlow* e o *Keras* API e foram treinados individualmente para cada conjunto de dados em uma GPU A100 com 40 Gigabytes de memória RAM do *Google Collaboratory Pro*. A mesma semente aleatória foi usada durante os experimentos. Os modelos passaram pelas mesmas etapas de treinamento, sendo avaliados no mesmo conjunto de testes. No capítulo 5, destacamos os modelos de AM que tiveram os melhores desempenhos nos experimentos em cada atividade proposta neste estudo.

5 RESULTADOS E DISCUSSÃO

Neste Capítulo apresentamos os resultados obtidos nos experimentos realizados em cada atividade proposta, destacamos os principais modelos preditivos de depressão na adolescência, os atributos que otimizam o desempenho dos modelos e discutimos sua correlação com a sintomatologia da depressão. Os resultados dos experimentos realizados nesta tese podem ser consultados no Apêndice C.

5.1 Previsão do humor deprimido atual

Nesta atividade, realizamos experimentos extensivos (unimodais e multimodais) em uma tarefa de classificação binária de depressão com base em dados dos *smartphones* dos adolescentes.

5.1.1 Experimentos unimodais

Nós executamos quatro experimentos unimodais com foco na previsão do escore binário de depressão do questionário MFQ: o primeiro utilizou somente atributos espectrais extraídos dos áudios de relatos, o segundo, somente atributos prosódicos extraídos dos áudios episódicos, o terceiro, somente atributos de atividade extraídos processados pelo EBM e por fim, utilizando somente atributos de mobilidade extraídos das coordenadas de GPS dos adolescentes.

As Figuras 18 e 19 comparam o desempenho dos algoritmos de aprendizado clássico e profundo na tarefa de classificação binária considerando a Acurácia Balanceada nos experimentos unimodais.

Os resultados indicam que o modelo preditivo com o melhor desempenho nos experimentos unimodais usando a agregação fixa, utilizou os atributos espectrais extraídos dos áudios de relatos associados a uma DNN. O modelo alcançou uma taxa de Acurácia Balanceada de 77,58% no conjunto de testes.

A estratégia adotada para os áudios episódicos falhou na agregação de dados em um período mais curto (fixa). O modelo que obteve o melhor desempenho utilizou uma 1DCNN e alcançou uma taxa de Acurácia Balanceada de 56,21%.

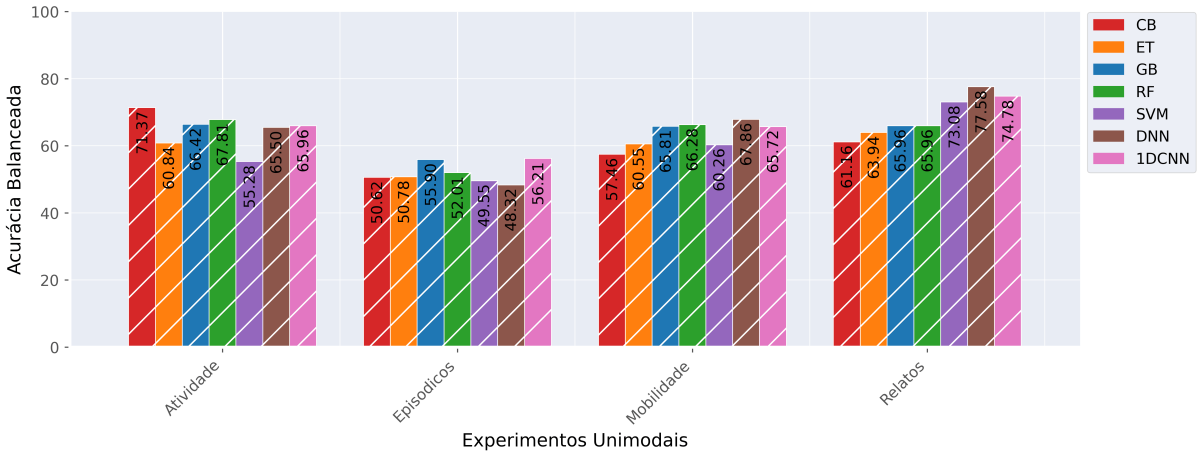


Figura 18 – Desempenho dos modelos unimodais usando agregação fixa

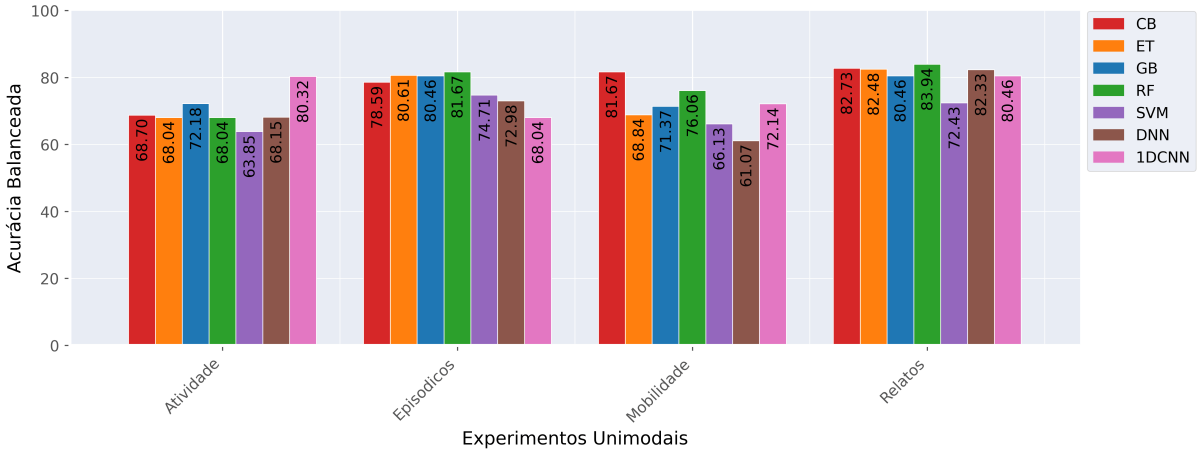


Figura 19 – Desempenho dos modelos unimodais usando agregação cumulativa

Os resultados evidenciam a dificuldade dos modelos em prever os escores binários de depressão no conjunto de dados usando agregação fixa, o que compromete a análise da importância dos atributos, especialmente considerando a distribuição assimétrica do conjunto de dados, onde aproximadamente 85% das amostras de adolescentes são da classe não deprimida.

No entanto, usando agregação cumulativa, a mesma estratégia obteve o segundo melhor desempenho médio entre os experimentos unimodais. Três classificadores obtiveram pelo menos 80% de Acurácia Balanceada no conjunto de testes, com destaque para o classificador RF que atingiu taxa de Acurácia Balanceada de 81,67%.

Será preciso rever a estratégia de extração de atributos adotada para os áudios episódicos, em pesquisas futuras, visto que para a abordagem fixa foram pouco representativos, enquanto na abordagem cumulativa obtiveram alta capacidade preditiva.

A extração de características espectrais dos áudios de relatos produziu o melhor resultado entre os experimentos unimodais e agregações. O modelo com melhor desempenho no experimento unimodal usando agregação cumulativa utilizou áudios de relatos, associados a um classificador RF e atingiu uma taxa de 83,94% de Acurácia

Balanceada no conjunto de testes.

Uma análise comparativa das agregações adotadas neste estudo, mostra que nenhum classificador conseguiu superar a barreira dos 80% de Acurácia Balanceada usando a agregação fixa. Ao mesmo tempo, podemos estabelecer modelos candidatos em todos os experimentos unimodais usando a agregação cumulativa. A Tabela 10 apresenta o resultado dos experimentos unimodais com maior detalhamento usando agregação cumulativa.

Tabela 10 – Experimentos unimodais realizados no conjunto de dados utilizando agregação cumulativa. Os melhores resultados obtidos pelos modelos preditivos de depressão são apresentados em negrito.

Experimentos Unimodais		Treinamento	Teste				
Modalidade	Classificador	B. Acc. média	Acc	Esp	Sen	B. Acc	F1
Áudios de Relatos	CB	87,92	89,73	92,74	**72,72**	82,73	68,08
	ET	95,92	**92,47**	96,77	68,17	82,48	**73,17**
	GB	89,37	89,03	92,74	68,17	80,46	65,22
	RF	94,47	91,78	95,16	**72,72**	**83,94**	72,72
	SVM	75,22	84,93	90,32	54,55	72,43	52,17
	DNN	85,06	89,03	91,94	**72,72**	82,33	66,67
	1DCNN	92,2	89,04	92,74	68,18	80,46	65,22
Áudios Episódicos	CB	85,85	89,03	93,55	63,63	78,59	63,63
	ET	96,48	**92,47**	**97,58**	63,63	80,61	71,78
	GB	84,98	89,03	92,74	68,17	80,46	65,22
	RF	93,7	91,10	95,16	68,17	81,67	69,77
	SVM	67,9	85,61	90,32	59,08	74,71	55,32
	DNN	81,78	89,03	95,97	50,0	72,98	57,89
	1DCNN	81,2	86,99	95,16	40,91	68,04	48,65
Atividade	CB	83,43	84,93	91,94	45,45	68,7	47,62
	ET	94,16	86,99	95,16	40,91	68,04	48,65
	GB	82,84	87,67	94,35	50,0	72,18	55,0
	RF	92,31	86,99	95,16	40,91	68,04	48,65
	SVM	66,93	76,71	82,26	45,45	63,85	37,04
	DNN	83,28	80,82	86,29	50,0	68,15	44,0
	1DCNN	86,8	85,61	87,9	**72,72**	80,32	60,38
Mobilidade	CB	78,79	91,10	95,16	68,17	81,67	69,77
	ET	94,27	88,36	96,77	40,91	68,84	51,43
	GB	84,44	86,3	92,74	50,0	71,37	52,38
	RF	91,09	91,10	**97,58**	54,55	76,06	64,86
	SVM	61,3	77,4	82,26	50,0	66,13	40,0
	DNN	77,53	81,51	90,32	31,81	61,07	34,15
	1DCNN	84,8	78,08	80,65	63,64	72,14	46,67

Nós consideramos que o experimento unimodal, que utilizou atributos espectrais extraídos dos áudios de relatos como dados de entrada, impactaram significativamente o desempenho dos classificadores.

O modelo preditivo com o melhor resultado no experimento unimodal baseado nos áudios de relatos usou classificador RF. Este modelo alcançou as maiores taxas de Acurácia Balanceada (83,94%) e Sensibilidade (72,72%), o que pode sugerir uma capacidade superior em diferenciar as amostras deprimidas de amostras não deprimidas dos adolescentes.

No experimento unimodal baseado em áudios episódicos, um classificador ET obteve as melhores taxas de Acurácia entre os experimentos unimodais superando os resultados do melhor experimento utilizando áudios de relatos em 0,69%, e aumentou a Especificidade em 2,42%. No entanto, ele retornou uma taxa de sensibilidade 9,09% inferior, o que sugere que este modelo enfrentou maiores dificuldades ao prever a classe minoritária (deprimida) do que o modelo preditivo usando áudios de relatos e classificador RF.

No experimento unimodal baseado nos dados de atividade, uma 1DCNN alcançou o melhor desempenho na tarefa de classificação. Esse modelo conseguiu prever os escores binários do sMFQ com taxa de Acurácia Balanceada de 80.32%, além de taxas de Acurácia, Especificidade e Sensibilidade de 85,61%, 87,9% e 72,72%, respectivamente.

No experimento unimodal usando dados de mobilidade para prever as pontuações binárias de depressão no sMFQ, o classificador CB superou outros modelos. Ele obteve valores de Acurácia Balanceada de 81,67%, juntamente com altas taxas de Acurácia (91,10)% e Especificidade (95,16%). Contudo, apresentou grande dificuldade para prever amostras da classe deprimida (taxa de Sensibilidade de 68,17%), problema também encontrado em experimentos usando áudios episódicos.

5.1.2 Experimentos multimodais

Nós realizamos 22 experimentos multimodais combinando as modalidades de dados utilizadas na etapa anterior e agregações para prever as pontuações binárias de depressão do questionário sMFQ.

A Figura 20 apresenta os resultados dos classificadores que obtiveram os melhores desempenhos para cada experimento na tarefa de classificação binária de depressão, considerando o tipo de agregação de dados e Acurácia Balanceada no conjunto de testes.

Os resultados sugerem que os modelos multimodais superam os unimodais na maioria dos experimentos para prever o estado deprimido atual dos adolescentes. Nossos achados indicam que a integração de múltiplas modalidades favoreceu o desempenho dos modelos preditivos, por permitirem captar características relevantes sobre o estado atual do humor deprimido dos adolescentes.

A combinação de modalidades com o melhor resultado foi alcançada, usando dados de atividade, áudios de relatos e mobilidade (experimento multimodal ARM). Espe-

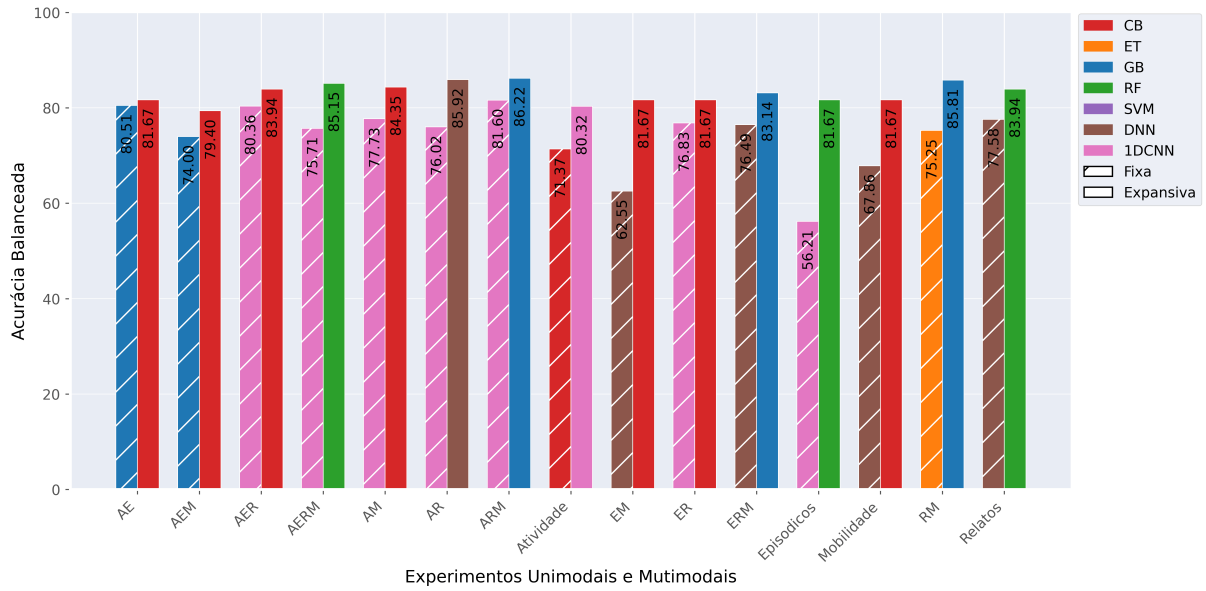


Figura 20 – Desempenho dos melhores modelos unimodais e multimodais usando agregação fixa e cumulativa

cificamente, uma DNN alcançou 81,60% de Acurácia Balanceada usando agregação fixa, enquanto um modelo baseado no classificador GB alcançou 86,22% de Acurácia Balanceada usando agregação cumulativa. Esses resultados evidenciam a importância de características de atividade, mobilidade e da fala dos adolescentes, coletadas em seu ambiente doméstico como indicadores de humor deprimido.

Existe uma grande variabilidade da frequência de coleta de dados do EBM, caracterizada pela alternância entre períodos (horas e até dias) de baixa coleta de dados, e momentos em que o EBM coletou uma quantidade expressiva de dados em um período mais curto de tempo.

Os resultados indicam que mudar a agregação de dados de uma janela fixa de dois dias para outra que acumula dados coletados em dias anteriores, proporcionou um ganho significativo de desempenho dos classificadores.

A agregação cumulativa mostrou ser mais tolerante à falta de dados do EBM e do IDEA-Bot, em comparação com a agregação fixa. Da mesma forma, permitiu aumentar o número de amostras válidas nos conjuntos de dados utilizados neste estudo, o que foi possível porque a agregação cumulativa pode diluir com eficiência a variabilidade da coleta de dados dos adolescentes do IDEA-Tech inerente ao processo de coleta passiva realizada pelo EBM.

Nossos achados sugerem que a agregação cumulativa é uma alternativa importante às estratégias de imputação de dados comumente adotadas em estudos que utilizam dados coletados passivamente para prever a depressão. A Tabela 11 apresenta com mais detalhes os modelos que obtiveram o melhor desempenho durante os experimentos multimodais usando agregação cumulativa.

O experimento multimodal AERM usou os atributos de todas as modalidades em

Tabela 11 – Resumo dos experimentos multimodais realizados utilizando agregação cumulativa. Os melhores resultados obtidos pelos modelos preditivos de depressão são apresentados em negrito.

Experimentos Multimodais		Treinamento	Testes				
Modalidade	Classificador	B. Acc média	Acc	Esp	Sen	B. Acc	F1
AE	CB	85,38	91,10	95,16	68,17	81,67	69,77
AEM	CB	86,59	90,41	95,16	63,63	79,4	66,67
AER	CB	89,39	91,78	95,16	72,72	83,94	72,72
AERM	RF	95,19	**93,84**	**97,58**	72,72	85,15	**78,05**
AM	CB	85,33	92,47	95,97	72,72	84,35	74,42
AR	DNN	87,03	85,61	85,48	**86,36**	85,92	64,41
ARM	GB	89,8	92,47	95,16	77,27	**86,22**	75,56
EM	CB	86,91	91,10	95,16	68,17	81,67	69,77
ER	CB	90,13	91,10	95,16	68,17	81,67	69,77
ERM	GB	90,55	90,41	93,55	72,72	83,14	69,57
RM	GB	89,82	91,78	94,35	77,27	85,81	73,91

conjunto. Como resultado, o classificador RF obteve maiores taxas de Acurácia (93,84%), Especificidade (97,58%) e medida F1 (78,05%) entre os experimentos, além de boa Acurácia Balanceada (85,15%). Contudo, uma taxa de Sensibilidade (72,72%) inferior indica que o modelo teve mais dificuldades para prever escore binário para depressão do sMFQ do que outros modelos candidatos.

O experimento multimodal AR usou atributos de atividade e áudio de relatos. Uma DNN obteve os melhores resultados, alcançando a maior taxa de Sensibilidade entre os experimentos (86,36%), um aumento de 9,09% em relação ao melhor experimento. Ao mesmo tempo, o modelo sofre uma perda de desempenho considerável em relação às demais medidas.

O experimento multimodal ARM usou atributos de atividade, áudios de relatos e mobilidade. O modelo que obteve o melhor desempenho foi associado ao classificador GB e alcançou taxas de Acurácia (92,47%), Especificidade (95,16%), Sensibilidade (77,27%) e medida F1 (75,56%). Este experimento obteve a maior taxa de Acurácia Balanceada (86,22%) entre os experimentos de previsão do humor deprimido atual.

Especificamente, obtivemos um ganho de desempenho de 4,62% de taxa de Acurácia Balanceada em comparação com o melhor modelo usando agregação fixa (também experimento ARM) e um aumento de 2,28% em relação ao melhor experimento unimodal que usa áudios de relatos e classificador RF.

A matriz de confusão apresentada na Figura 21 permite visualizar o desempenho do classificador GB para todos os valores reais das amostras e previstos pelo modelo de classificação binária de depressão. Os resultados sugerem que este modelo é mais equilibrado entre as taxas de Especificidade e Sensibilidade, o que o torna mais eficiente para discriminar amostras deprimidas e não deprimidas dos adolescentes.

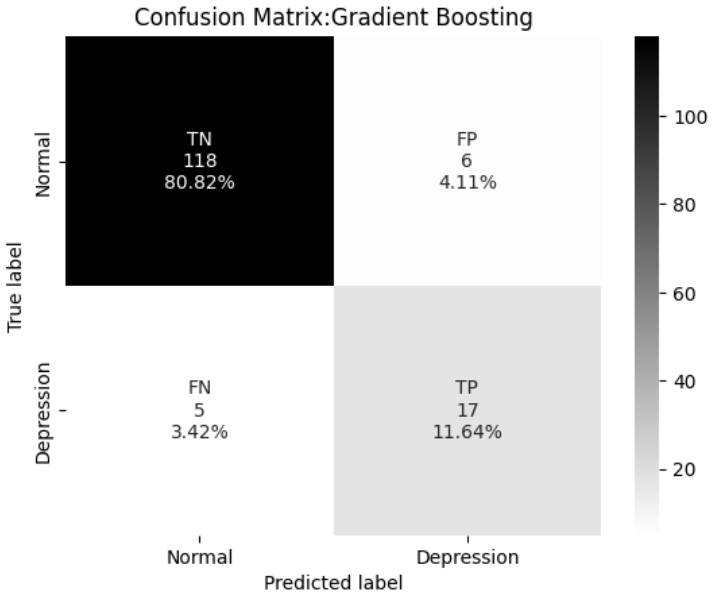


Figura 21 – Experimento Multimodal ARM: Matriz de confusão

Nós apresentamos na Tabela 12, o desempenho do modelo preditivo no experimento ARM com classificador GB e agregação cumulativa para as amostras masculinas e femininas do nosso conjunto de testes. É importante ressaltar que os modelos desenvolvidos neste estudo não foram treinados com a informação do sexo dos adolescentes, uma vez que esse não era o foco principal da pesquisa.

Tabela 12 – Desempenho do modelo preditivo de humor deprimido atual para ambos os sexos.

Modalidade	Classificador	Sexo	Acc	Spe	Sen	B. Acc	F1
ARM	GB	Masc	98,48	95,16	0	50,0	0
		Fem	87,50	95,16	80,95	85,39	77,72

O modelo apresentou alta Acurácia para o sexo masculino (98,48%), entretanto obteve um desempenho muito ruim na detecção de amostras deprimidas entre adolescentes do sexo masculino, principalmente devido à taxa de sensibilidade, que se mostrou uma grande limitação para a previsão das amostras deprimidos neste grupo.

Por outro lado, o modelo apresentou desempenho consideravelmente melhor para classificar a depressão entre adolescentes do sexo feminino, alcançando uma taxa de sensibilidade de 80,95%. Além disso, o modelo atingiu uma taxa de Acurácia Balanceada de 85,39% para o mesmo grupo.

A principal limitação deste experimento decorre da divisão estratificada dos dados, que resultou em um conjunto de testes formado por 146 amostras, das quais 22 eram deprimidas (sendo 21 de mulheres e apenas 1 de homens). O processo de estratificação dos dados prejudicou a análise do desempenho do modelo na previsão de casos de depressão no grupo masculino, uma vez que esse grupo representa apenas 2% das amostras do conjunto de dados.

As pesquisas de Jiang et al. (2017); Liu et al. (2020) e Bailey; Plumbley (2020) exploram um desafio comum em modelos de detecção de depressão, o viés dos modelos em relação ao sexo dos indivíduos. Estes estudos discutem como um tratamento amostral baseado nesta variável pode mitigar seus efeitos e melhorar o desempenho dos classificadores.

Neste contexto, é de suma importância identificar e reduzir o viés de sexo em nossos experimentos. Acreditamos que para melhorar o desempenho do modelo preditivo no grupo masculino pode ser necessário obter mais dados de adolescentes do sexo masculino com depressão ou considerar abordagens específicas de sobreamostragem para tratar o desequilíbrio dos conjuntos de dados com base no sexo dos adolescentes.

Nós consideramos o experimento multimodal ARM associado a um Classificador GB e agregação cumulativa o modelo com o melhor desempenho entre os experimentos para prever as pontuações binárias do sMFQ. Assim, adotamos o algoritmo de SHAP para identificar a contribuição dos atributos para previsão do modelo.

5.1.3 Análise de importância dos atributos usados para previsão do humor deprimido atual

Calculamos a importância dos atributos do modelo com o melhor desempenho selecionado em experimentos unimodais e multimodais. Os resultados são apresentados por meio de gráficos de importância global e de direcionalidade para destacar os 20 principais atributos que mais influenciaram a previsão do modelo nas amostras do conjunto de testes.

A combinação de dados de atividade, de áudios de relatos e mobilidade associada a um classificador GB produziu o modelo com maior capacidade de prever o humor deprimido dos 93 adolescentes do IDEA-Tech usados neste estudo. Identificamos algumas características derivadas desses dados que podem apresentar uma correlação significativa com os sintomas da depressão.

O gráfico da importância global do experimento multimodal ARM (Figura 22) aponta para uma maior contribuição dos atributos espectrais na saída do modelo preditivo. Especificamente, MFCC, espectrogramas, fluxo, curtose, contraste e ruído espectral.

Os resultados indicam que os coeficientes MFCC *accelerate* (3, 16, 17, 19, 20, 27, 28), espectrogramas log-mel (1, 3, 4, 13, 27, 124) e contraste (0), estão entre os mais importantes para prever os escores binários da depressão do sMFQ (Figura 23).

A intensidade vocal, frequência de falas e pausas, afinação são frequentemente afetadas pela depressão. Por esse motivo, é comum que os pacientes apresentem discursos reduzidos e mais espaçados, baixa velocidade e intensidade da fala, monotonicidade do tom de voz, entre outras alterações da fala (CUMMINS et al., 2015; STASAK et al., 2016; REJAIBI et al., 2019).

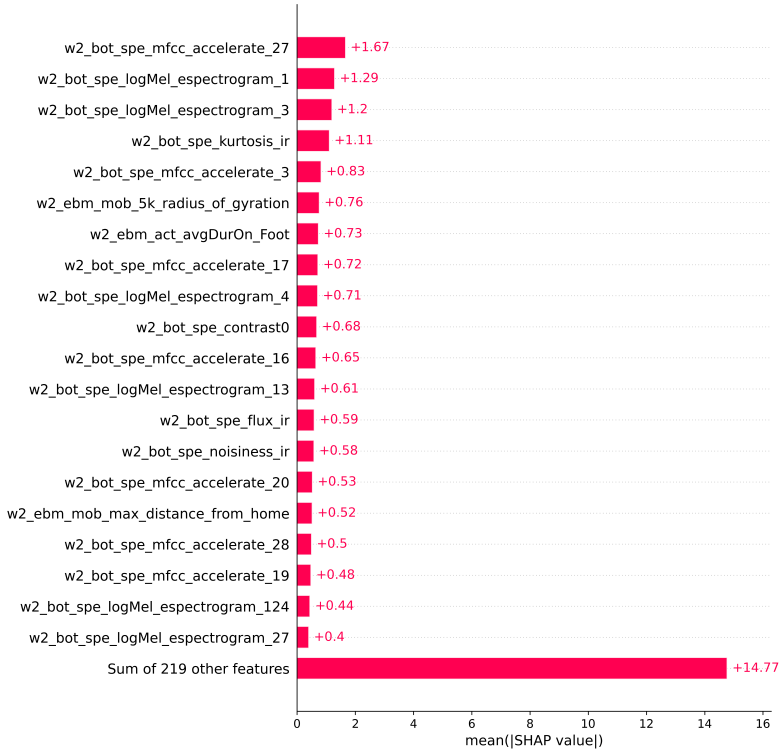


Figura 22 – Experimento Multimodal ARM: Importância global

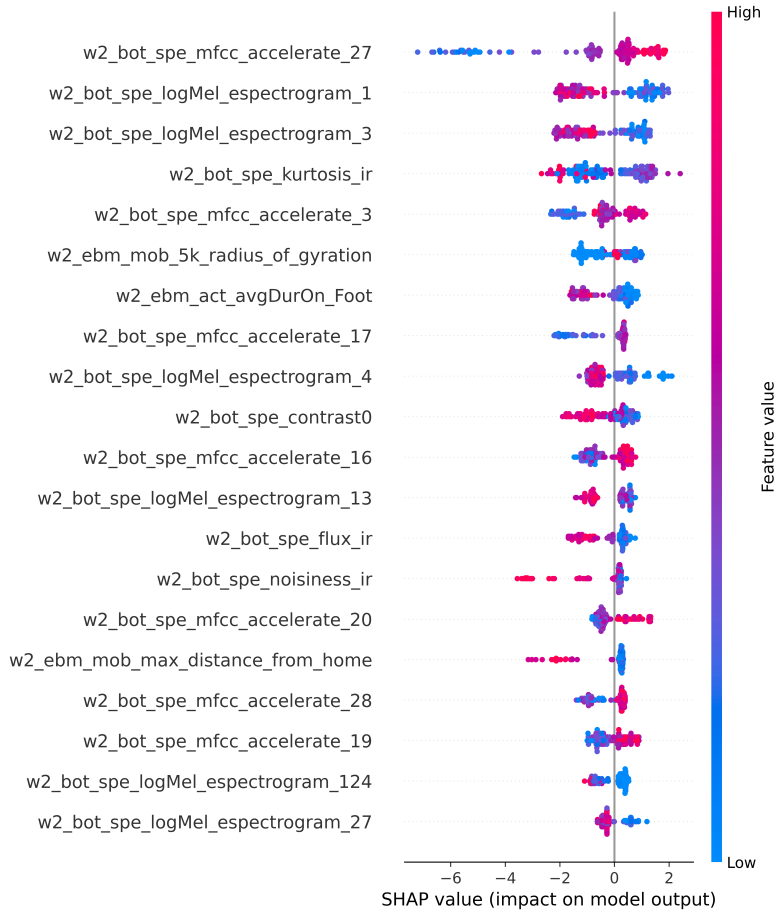


Figura 23 – Experimento Multimodal ARM: Direcionalidade

Os valores do coeficiente MFCC, apresentados na Figura 24, foram superiores nas amostras deprimidas em comparação às amostras não deprimidas do sMFQ, com destaque para os coeficientes 3 e 17. Por outro lado, os valores dos coeficientes do espectrograma log-mel nas amostras deprimidas foram menores do que nas amostras não deprimidas do sMFQ, especialmente coeficientes 1 e 3. Encontramos o mesmo padrão para o primeiro coeficiente do contraste espectral (0) mas com menos influência na previsão do modelo.

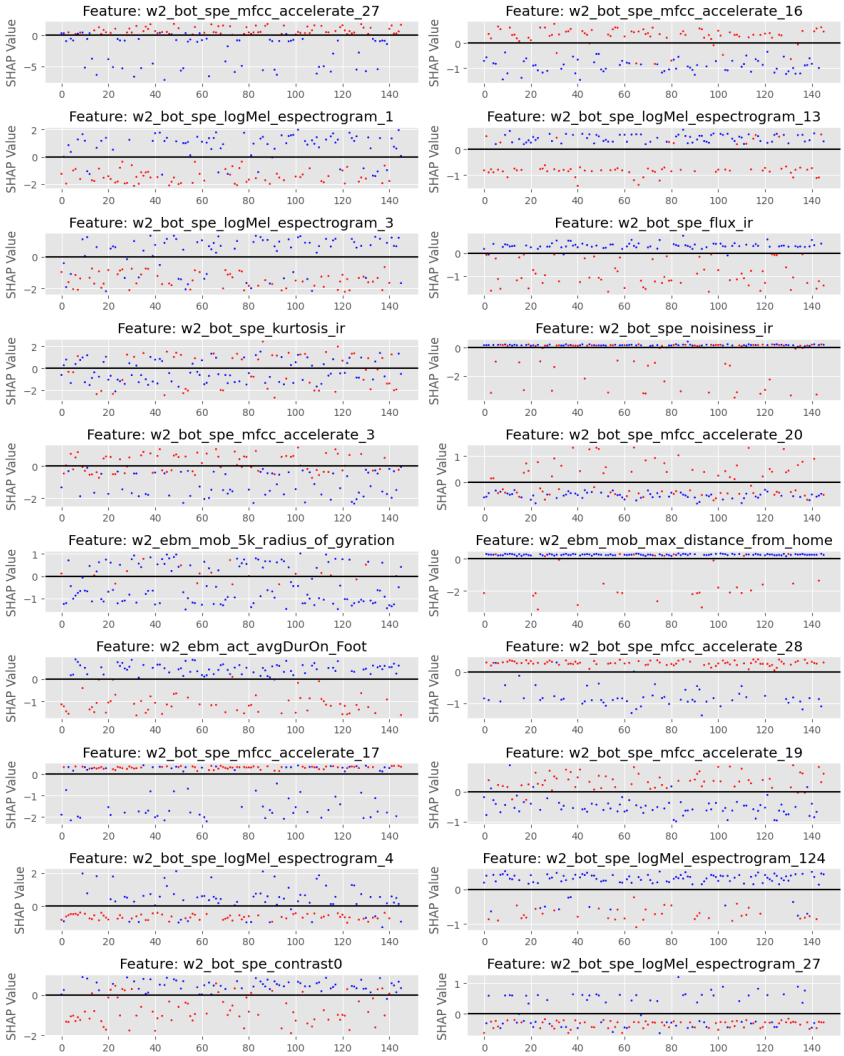


Figura 24 – Experimento Multimodal ARM: Direcionalidade por atributo

Nossos achados corroboram estudos anteriores que investigaram os coeficientes espectrais como marcadores digitais da fala deprimida. Estes estudos encontraram associações entre os coeficientes espectrais e a presença de sintomas da depressão e demência (SUMALI et al., 2020; KIM et al., 2023) e risco de suicídio na fala (MIN et al., 2023).

Além disso, estudos recentes sugerem que os coeficientes MFCC podem estar associados a um trato vocal mais estreito em pacientes deprimidos, possivelmente causado pelo retardo psicomotor. Coeficientes MFCC se mostraram mais altos em

pacientes deprimidos e são minimamente afetados por características sociodemográficas como idade e sexo conforme análises conduzidas por Taguchi et al. (2018) e Zhao et al. (2022). Por outro lado, a pesquisa de Wang et al. (2023) revelou que valores de coeficientes MFCC *accelerate* diminuíram conforme os pacientes deprimidos submetidos a intervenção medicamentosa, apresentavam melhora.

Os coeficientes MFCC *accelerate* e espectrogramas log-mel tiveram um impacto significativamente maior na previsão do modelo. Porém, é importante ressaltar, que eles representam a energia em diferentes bandas de frequências de um sinal. Em outras palavras, eles são fragmentos de um conjunto maior de características que formam o atributo espectral. O contraste, MFCC e espectrograma adotados neste estudo utilizaram 7, 40 e 128 coeficientes espectrais para representar um sinal da fala, respectivamente.

Adicionalmente, identificamos uma associação positiva entre menores níveis de ruído e fluxo espectral e o escore deprimido do sMFQ. Esses achados sugerem que a presença de ruído e fluxo espectral reduzidos podem estar associadas a sinais de fala deprimida dos adolescentes.

Muitas pesquisas usaram atributos espectrais para reconhecimento de fala devido à sua robustez na descrição do sinal em frequências mais baixas. Os MFCC (REJAIBI et al., 2019; SUWANNAKHUN; YINGTHAWORNSUK, 2019; ALGHIFARI et al., 2019) e espectrogramas (DINKEL et al., 2019; LIN et al., 2020) foram adotados com eficiência em estudos de detecção de depressão quando comparados a outros LLDs, isoladamente.

Em nosso estudo, as características espectrais extraídas dos áudios de relatos exerceram uma influência significativa sobre os modelos preditivos de depressão. Contudo, seu desempenho foi melhorado quando foram usadas em conjunto com características de atividade e mobilidade dos adolescentes.

O atributo de atividade, duração média de caminhadas (AvgDurOnFoot) e os atributos de mobilidade, raio de giro formado pelos cinco locais visitados com maior frequência pelos adolescentes (5k_radius_of_gyration) e a distância máxima de casa (mob_max_distance_from_home) completam os 20 atributos que mais influenciaram a predição do modelo de detecção de depressão, avaliados no conjunto de testes.

Diversos estudos que investigaram as características de mobilidade e atividade como marcadores de depressão sugerem que uma maior mobilidade e a prática de atividades físicas regulares podem estar associadas ao bem-estar do paciente. Em contrapartida, uma mobilidade reduzida ou uma rotina limitada de atividades físicas podem ser observados na sintomatologia da depressão (MASUD et al., 2020; NGUYEN et al., 2021; TØNNING et al., 2021).

Nós identificamos uma associação positiva entre um tempo médio de caminhada aumentado e uma maior distância percorrida a partir de casa e as amostras não-

deprimidas dos adolescentes. Estes resultados podem ser explicados, pois indivíduos deprimidos tendem a se movimentar menos pelo espaço geográfico que indivíduos saudáveis, conforme mencionado no trabalho de Moshe et al. (2021).

Nossos resultados estão alinhados com a literatura recente sobre a depressão. Os resultados do estudo de Tønning et al. (2021) indicaram que pontuações mais altas do HDRS-17 e no FAST estão associadas ao menor humor e atividade física reduzida relatadas pelos pacientes para os sintomas da depressão. Em Velazquez et al. (2022), os adolescentes que relataram pontuações mais baixas no PACE+ apresentaram pontuações mais altas no PHQ-A, indicando que adolescentes com níveis mais baixos de atividade física ou que não se exercitam com frequência podem ter mais sintomas da depressão. Portanto, nossos achados sugerem que a atividade física e a mobilidade podem ter um efeito preventivo contra a depressão em adolescentes.

Apesar da contribuição dada ao desempenho do modelo preditivo, os gráficos de importância não mostraram uma direção clara para os atributos de raio de giro e a curtose espectral.

Assim, os atributos de atividade, da fala e da mobilidade dos adolescentes destacados neste estudo revelaram características importantes para investigar os sintomas da depressão, não observáveis sob a perspectiva da avaliação tradicional baseada em entrevistas clínicas e questionários autorrelatados.

5.2 Previsão do humor deprimido futuro

Nós realizamos diversos experimentos de classificação binária e regressão para avaliar o desempenho dos modelos nomotéticos e idiográficos de detecção de depressão desenvolvidos neste estudo.

5.2.1 Previsão de amostras da próxima semana

Esta atividade compara modelos de AM para detectar o humor deprimido futuro usando as pontuações do sMFQ coletados em semanas diferentes. Especificamente, usamos as amostras coletadas entre D1 a D8 no conjunto de treinamento, e as amostras coletadas entre D9 a D14 no conjunto de teste para avaliar os modelos preditivos.

5.2.1.1 Experimentos nomotéticos e idiográficos

Para a abordagem nomotética, conduzimos experimentos monotarefas e multitarefa que envolvem classificação binária e regressão. Os experimentos monotarefa incluem modelos de aprendizado clássico (CB, ET, GB, RF, SVM) e modelos de aprendizado profundo (DNN, 1DCNN). Eles usaram o método de busca aleatória de hiperparâmetros com 10 iterações usando o *RandomizedSearchCV*. Os experimentos multitarefa, adicionam um modelo de aprendizado profundo (MT1DCNN). Os modelos

multitarefa usaram otimização Bayesiana de hiperparâmetros e realizaram 20 interações usando algoritmo TPE.

Para abordagem idiográfica, usamos um modelo MT1DCNN associado a matrizes de similaridade calculadas com base na similaridade do cosseno (MT1DCNN-C), distância euclidiana (MT1DCNN-E) e similaridade de *Jaccard* (MT1DCNN-J).

A Figura 25 apresenta um comparativo entre modelos preditivos de depressão usando abordagens nomotéticas e idiográficas em diferentes agregações para prever os escores binários de depressão do sMFQ considerando a Acurácia Balanceada no conjunto de testes.

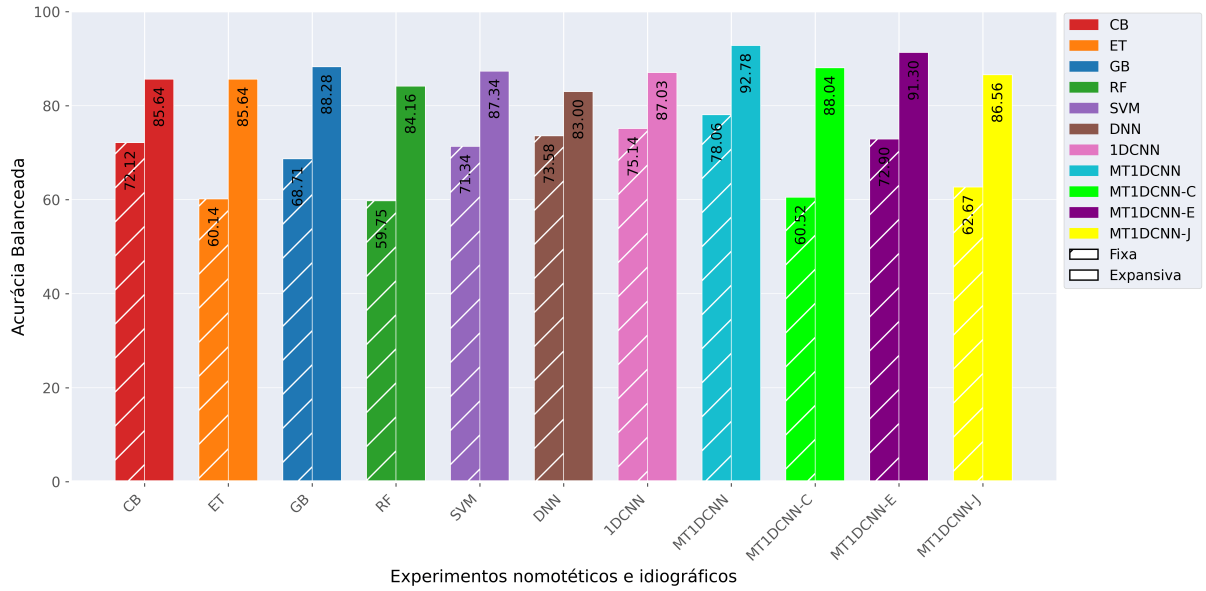


Figura 25 – Desempenho dos modelos nomotéticos e idiográficos na tarefa de classificação binária. As taxas de Acurácia Balanceada são obtidas na previsão dos questionários sMFQ no conjunto de testes.

Os resultados indicam que os modelos nomotéticos obtiveram um desempenho pior quando usam agregação fixa, enquanto, apresentaram melhoras significativas usando agregação cumulativa. O classificador GB usando agregação cumulativa foi o modelo monotarefa que obteve o melhor resultado na tarefa de classificação binária. Ele atingiu uma taxa de Acurácia Balanceada de 88,28%.

O modelo multitarefa MT1DCNN alcançou o melhor desempenho entre os modelos nomotéticos tanto na agregação fixa quanto na agregação cumulativa, atingindo taxas de Acurácia Balanceada de 78,06% e 92,78%, respectivamente. Esse resultado representa um ganho de 14,72% na taxa de Acurácia Balanceada, apenas mudando como os dados foram agregados ao longo do tempo para calcular os atributos de atividade, relatos e mobilidade, utilizados como entrada do modelo preditivo.

Os resultados também sugerem que o desempenho dos modelos idiográficos diminui quando usam agregação fixa. O modelo idiográfico MT1DCNN-E alcançou 72,90% de Acurácia Balanceada e produziu o melhor modelo idiográfico usando agregação

fixa. O desempenho do modelo foi 5,16% inferior, se comparado ao melhor modelo nomotético (MT1DCNN) usando a mesma agregação.

Os modelos idiográficos obtiveram alta capacidade preditiva usando agregações de dados cumulativa. O modelo idiográfico MT1DCNN-E usando agregação cumulativa produziu o melhor modelo personalizado para prever as amostras de sMFQ futuras, atingindo uma taxa de Acurácia Balanceada de 91,30%, no entanto, ele foi superado em 1,48% pelo modelo MT1DCNN nomotético.

A Figura 26 apresenta o desempenho dos modelos nomotéticos e idiográficos para prever a intensidade dos sintomas da depressão em diferentes agregações considerando as taxas de erro médio e raiz do erro médio quadrático.

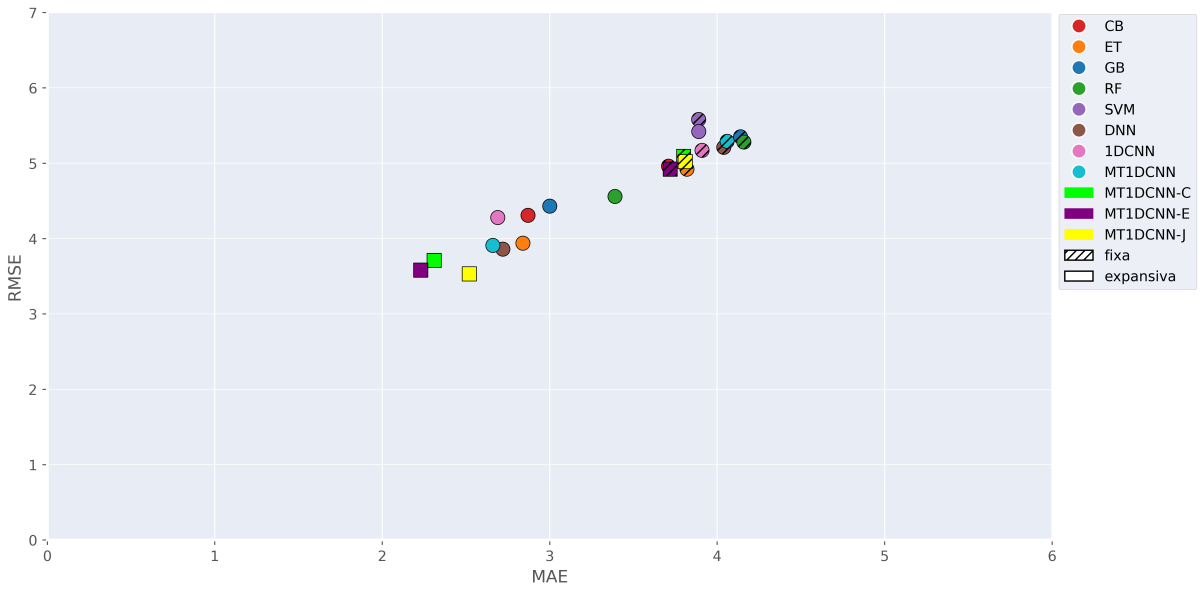


Figura 26 – Desempenho dos modelos nomotéticos e idiográficos na tarefa de regressão. As taxas de erro são obtidas na previsão dos sMFQ no conjunto de testes.

Os resultados dos experimentos na tarefa de regressão indicam uma melhora nas taxas de erro dos modelos que usaram agregação cumulativa. Uma 1DCNN usando agregação cumulativa produziu o melhor modelo monotarefa com valores de MAE e RMSE de 2,69 e 4,01, respectivamente, enquanto o modelo MT1DCNN alcançou taxas de MAE (2,66) e RMSE (3,91), ligeiramente melhores. No entanto, os modelos idiográficos conseguiram alcançar taxas MAE e RMSE reduzidas se comparados aos modelos nomotéticos.

Os resultados sugerem que os modelos idiográficos foram mais eficientes em reduzir os erros de previsão da intensidade dos sintomas da depressão usando agregação cumulativa, com destaque para o modelo idiográfico (MT1DCNN-E) que obteve o melhor desempenho na tarefa de regressão considerando as duas medidas de erro.

O MT1DCNN-E pôde prever a pontuação total do sMFQ com uma taxa de erro de 2,23 pontos em média considerando o intervalo da pontuação do sMFQ respondidos pelos adolescentes que vai de 0 a 23 em nossos conjuntos de dados. Isso representa

uma taxa de erro médio relativo de 10% na previsão da pontuação total do sMFQ. No entanto, o alto valor de RMSE indica que alguns erros de predição (cerca de 15%) podem ser significativamente maiores.

A Tabela 13 resume as medidas alcançadas pelos modelos que obtiveram melhores resultados nas tarefas de classificação binária e regressão usando agregação cumulativa.

Tabela 13 – Melhores experimentos nomotéticos e idiográficos usando agregação cumulativa. Os melhores resultados obtidos pelos modelos preditivos são apresentados em negrito.

Abordagem	Classificador	Acc	Esp	Sen	B. Acc	F1	MAE	RMSE
Nomotética	GB	92,59	94,41	82,14	88,28	76,67	-	-
Nomotética	1DCNN	-	-	-	-	-	2,69	4,28
Nomotética	MT1DCNN	**95,23**	96,27	**89,28**	**92,77**	**84,74**	2,66	3,91
Idiográfica	MT1DCNN-E	**95,23**	**96,89**	85,71	91,30	84,21	**2,23**	**3,58**

Os resultados sugerem que experimentos multitarefas usando otimização bayesiana foram muito eficazes para prever o humor deprimido futuro dos adolescentes do IDEA-Tech. Além disso, a agregação cumulativa melhorou o desempenho dos modelos de forma geral usando ambas as abordagens.

Nós consideramos que o modelo MT1DCNN com agregação cumulativa obteve o melhor desempenho entre os experimentos para prever o humor deprimido futuro em amostras semanais. Este modelo alcançou os maiores valores de Acurácia (95,25%), Sensibilidade (89,28%), Acurácia Balanceada (92,77%) e F1 (84,74%), além de uma boa capacidade de diferenciar amostras deprimidas e não deprimidas representada pelo equilíbrio entre as medidas de Especificidade e Sensibilidade.

Além disso, nós apresentamos na Tabela 14, o desempenho do modelo nomotético MT1DCNN para prever os escores da depressão do sMFQ em ambos os sexos, separadamente. A divisão semanal dos dados resultou em um conjunto de testes formado por 189 amostras, das quais 28 eram deprimidas (sendo 21 de mulheres e apenas 7 de homens).

Tabela 14 – Desempenho do modelo preditivo de humor deprimido futuro para ambos os sexos.

Classificador	Sexo	Acc	Spe	Sen	B. Acc	F1	MAE	RMSE
MT1DCNN	Masc	96,42	98,70	71,42	85,06	76,92	2,33	3,67
	Fem	94,28	94,04	95,23	94,64	86,95	2,92	4,09

Embora nossos conjuntos de dados sejam bastante desequilibrados, o modelo MT1DCNN obteve altas taxas de Acurácia e especificidades para ambos os sexos. Entretanto, um valor de 71,42% de sensibilidade indica um desempenho do modelo bastante inferior para identificar amostras deprimidas masculinas se comparado a taxa de sensibilidade obtida pelo modelo no grupo feminino que foi de 95,64%.

É importante observar que o modelo obteve um desempenho bastante robusto de uma forma geral, alcançando taxas de Acurácia superiores a 95% e Acurácias balanceadas significativas para ambos os sexos (85,06% para homens e 94,64% para mulheres). Portanto, nossos resultados apontam que o modelo pode ser eficaz para identificar a depressão em ambos os sexos, com um desempenho superior em amostras do sexo feminino.

Nós hipotetizamos que o desempenho do modelo MT1DCNN pode ter sido influenciado pela quantidade de amostras usadas durante o treinamento. Além disso, características intrínsecas dos atributos podem ter contribuído para identificar a maioria dos casos de depressão entre as adolescentes do sexo feminino.

Esses achados convergem com a literatura recente, os estudos conduzidos por Bailey; Plumbley (2020) e Liu et al. (2020) que indicam que modelos baseados em AM apresentam um desempenho superior ao prever a depressão em mulheres, mas devem ser validados em outras populações e pesquisas sobre a depressão.

A matriz de confusão (Figura 27) apresenta as predições do modelo para as classes deprimidas e não deprimidas no conjunto de testes.

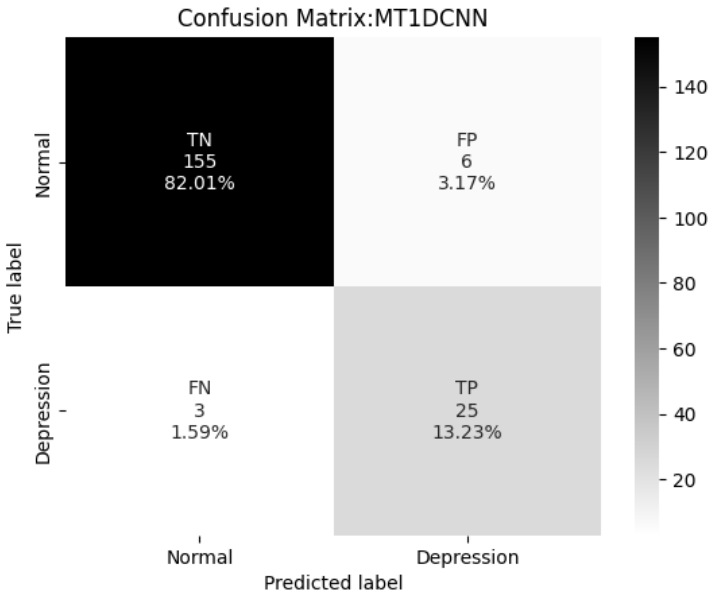


Figura 27 – Experimento Multimodal ARM - Previsão da próxima semana: Matriz de confusão

A seguir, nós avaliamos a importância relativa dos 20 principais atributos que mais contribuíram para melhorar o desempenho do modelo nomotético MT1DCNN.

5.2.1.2 Análise de importância dos atributos usados para previsão do humor deprimido futuro

O gráfico de importância do modelo de predição de depressão MT1DCNN apresentado na Figura 28 aponta para as contribuições dos atributos de atividade, *On_Foot, Tilting, In_Vehicle, Still e Unknown*.

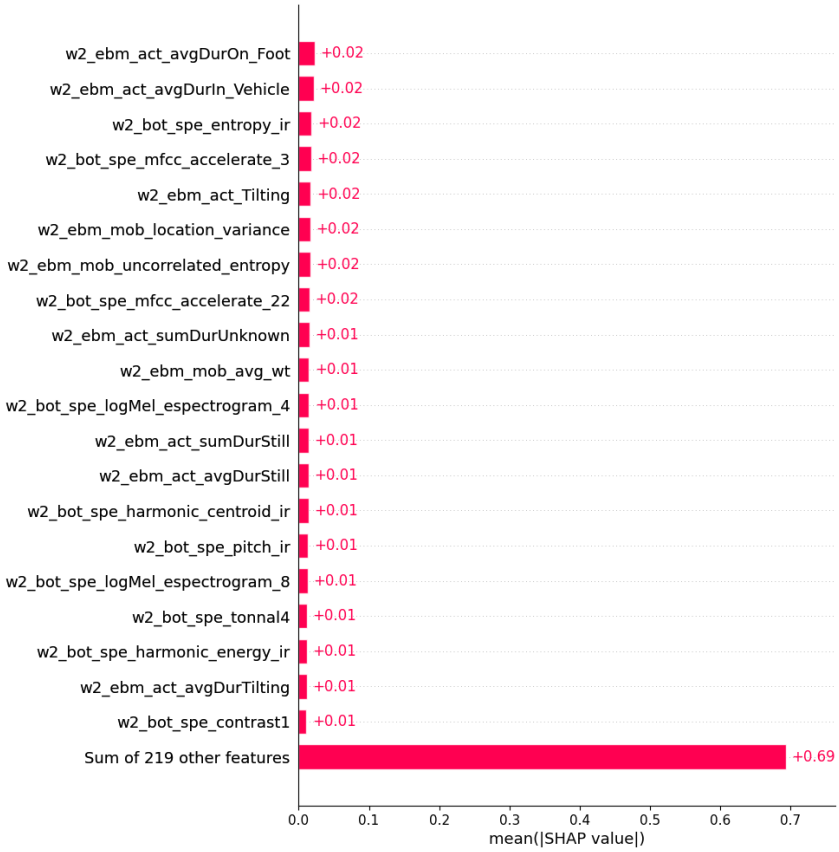


Figura 28 – Experimento Multimodal ARM - Previsão da próxima semana: Importância global

A análise de dados de sensores acelerômetros coletados em ambientes domésticos oportuniza o monitoramento dos níveis de atividade física de pacientes deprimidos em suas atividades cotidianas.

Nós podemos destacar as contribuições das atividades *On_Foot* e *Tilting* como principais indicadores de humor deprimido. Os resultados apontaram uma associação positiva entre um tempo médio de caminhada diária reduzido (avgDurOn_Foot) e as amostras deprimidas dos adolescentes. Esses resultados podem indicar manifestações de depressão quando interpretados como sinais de atividade física reduzida ocasionada por anedonia ou baixa energia, como sugerem os estudos de Pedrelli et al. (2020) e Masud et al. (2020).

Além disso, uma maior frequência em que um adolescente (*smartphone*) se move, caracterizada pelo atributo de atividade *act_Tilting* também foi associado à classe deprimida do sMFQ em nosso experimento (Figura 29).

Diversos atributos espectrais também contribuíram significativamente para o desempenho do modelo preditivo. No entanto, não observamos a mesma predominância dos coeficientes MFCC, e espectrogramas entre os 20 mais relevantes para a previsão do humor deprimido futuro, como ocorreu na atividade anterior. Este fato pode sugerir que tais atributos foram mais eficazes ao prever o humor deprimido atual dos adolescentes.

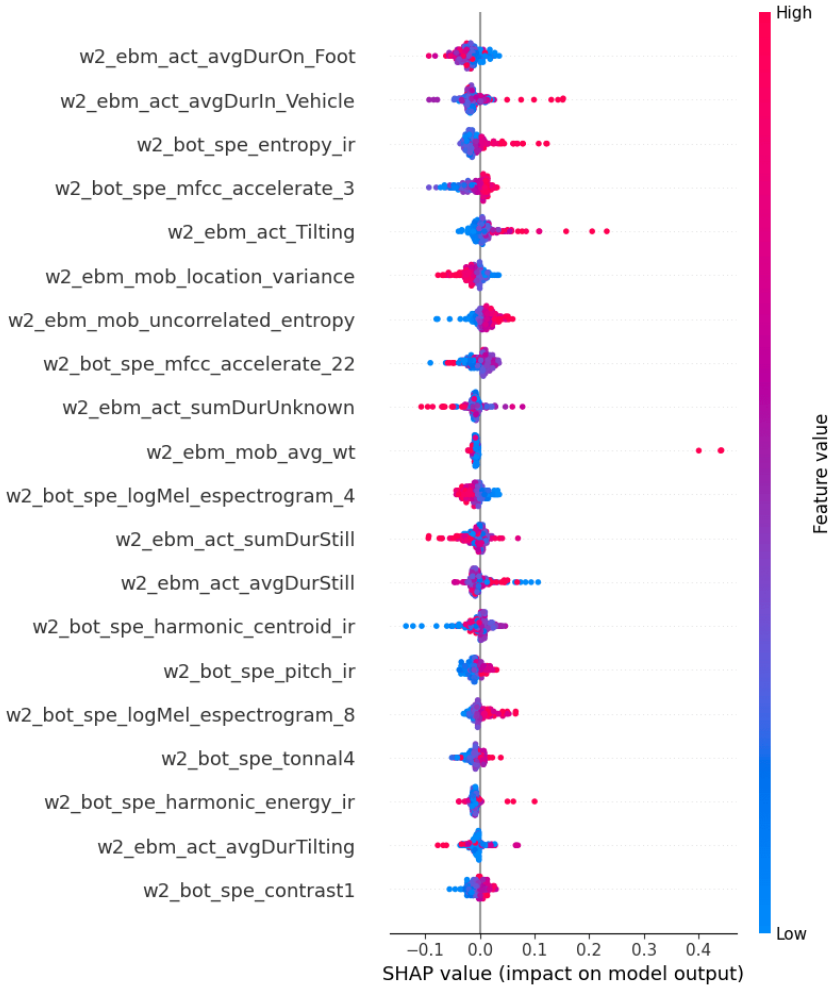


Figura 29 – Experimento Multimodal ARM - Previsão da próxima semana: Direcionalidade

Nós identificamos os coeficientes espectrais, MFCC, espectrogramas, contraste e a tonalidade espectral. Além do *pitch*, centróide harmônico e entropia espectral estão entre os atributos espectrais mais importantes para prever os escores binários da depressão do sMFQ.

Novamente, o coeficiente MFCC_accelerate (3), mostrado no gráfico de importância (Figura 30) se mostrou mais alto nas amostras deprimidas do que nas amostras não deprimidas dos adolescentes. Reconhecemos padrão semelhante para o segundo coeficiente de contraste espectral (1) e para o coeficiente de tonalidade (4).

Além disso, os resultados indicam uma associação positiva entre taxas mais altas de *pitch* e da entropia espectral e o escore deprimido do sMFQ. No entanto, identificamos uma direcionalidade divergente entre os coeficientes do espectrograma log-mel (4 e 8).

Nós podemos destacar os atributos variação de localização e entropia não correlacionada como importantes indicadores de depressão derivados da mobilidade dos adolescentes. Eles integram os 20 principais atributos que mais influenciaram a previsão do modelo de detecção de depressão.

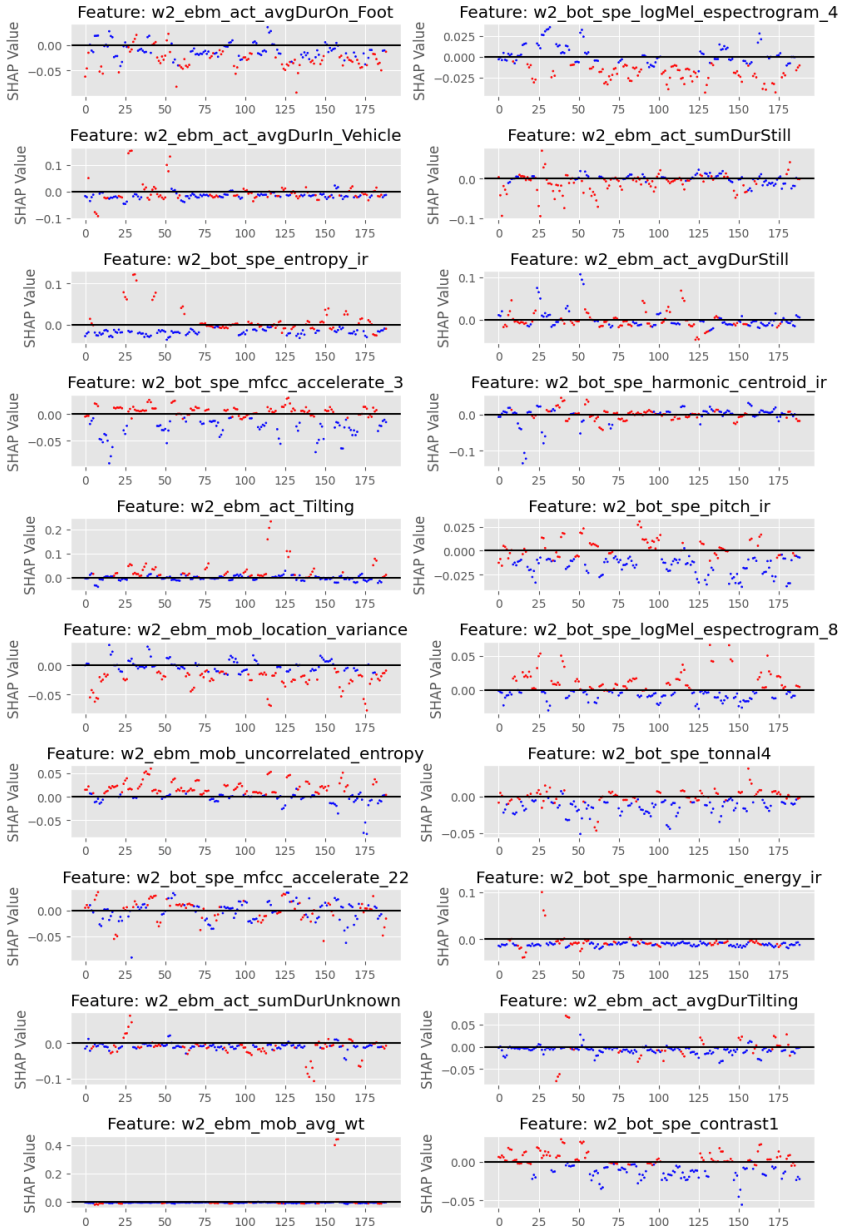


Figura 30 – Experimento Multimodal ARM - Previsão da próxima semana: Direcionalidade por atributo

Nosso estudo revelou que o escore binário de depressão do sMFQ correlacionou-se fortemente com uma variação de localização reduzida (*location_variance*) dos adolescentes e um aumento de sua entropia não correlacionada (*uncorrelated_entropy*).

Ambos os atributos medem a movimentação dos adolescentes usando as coordenadas de GPS. Em síntese, a variação de localização calcula a variação dos locais visitados pelos adolescentes, enquanto a entropia não correlacionada mede o nível de previsibilidade do movimento do adolescente considerando a frequência em que locais foram visitados anteriormente (SONG et al., 2010; PAUL et al., 2017; MOSHE et al., 2021).

Nossas observações reforçam as descobertas de pesquisas anteriores, as quais

demonstraram que essas características da mobilidade do indivíduo podem ser importantes indicadores de depressão.

Saeb et al. (2015, 2016) avaliaram vários atributos extraídos de dados de localização para prever a depressão. Ambos os trabalhos concluíram que a variação da localização, a entropia normalizada e o movimento circadiano foram os atributos de localização mais fortemente correlacionados com os sintomas da depressão.

Nossos resultados também convergem parcialmente com as pesquisas desenvolvidas por Masud et al. (2020) e Moshe et al. (2021). Ambos os estudos, constataram que uma menor diversidade de locais visitados pode estar associada a gravidade da depressão, sendo percebida por valores de variação de localização reduzidos.

Embora tenham contribuído para o desempenho do modelo preditivo, não foi possível determinar um direcionamento objetivo para os atributos de atividade, (*sumDurUnknown, sumDurStill, avgDurTilting*), dos áudios de relatos (centróide e energia harmônicos) e de mobilidade, tempo de deslocamento (avg_wt).

5.2.2 Previsão dos próximos sMFQ

A segunda atividade de previsão do humor deprimido futuro, está centrada na previsão dos próximos questionários sMFQ respondidos pelos adolescentes do IDEA-RiSCo.

5.2.2.1 Experimentos nomotéticos e idiográficos

Durante o ciclo completo de coleta de dados de 14 dias (D1 a D14), os adolescentes responderam sete questionários sMFQ com frequência de uma vez a cada 2 dias. Os dados obtidos nos primeiros dois dias (D1 e D2) foram usados exclusivamente no treinamento dos modelos preditivos. A partir daí, os dias de coleta subsequentes foram incorporados a cada dobra durante o treinamento para a previsão dos próximos questionários sMFQ.

Nós conduzimos quatro experimentos multimodais focados em prever os próximos questionários sMFQ. A Figura 31 ilustra o desempenho dos modelos MT1DCNN e MT1DCNN-E para discriminar amostras deprimidas e não deprimidas, considerando a Acurácia Balanceada média, nos dois métodos de agregação.

As taxas de Acurácia Balanceada melhoram gradualmente à medida que mais informações anteriores foram adicionadas ao treinamento dos modelos preditivos. Entretanto, as taxas de Acurácia Balanceada obtidas pelos modelos preditivos nos experimentos que utilizaram agregação fixa foram bastante reduzidas.

Os resultados evidenciam a dificuldade dos modelos em prever os escores binários da depressão do sMFQ usando a agregação de dados fixa, enquanto, os experimentos que usaram agregação de dados cumulativa obtiveram resultados superiores para ambos os modelos preditivos.

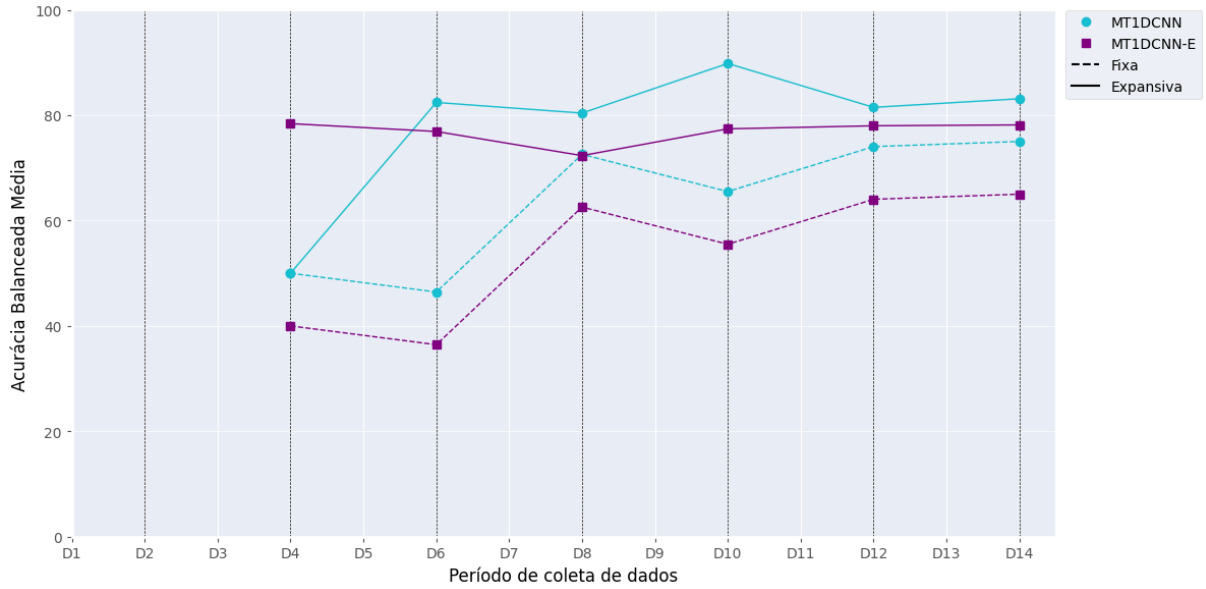


Figura 31 – Desempenho dos modelos nomotético (MT1DCNN) e idiográfico (MT1DCNN-E) na tarefa de classificação binária dos próximos sMFQ. A linha vertical indica o dia em que os sMFQ foram coletados. Os dados correspondentes ao primeiro sMFQ são usados exclusivamente no treinamento do modelo. As taxas de Acurácia Balanceada são obtidas pela previsão do sMFQ em cada dobra do conjunto de validação.

Nossas descobertas indicam que o modelo MT1DCNN-E demonstrou maior estabilidade ao longo do tempo, resultando em taxas de Acurácia Balanceada mais consistentes na previsão dos sMFQ. Acreditamos que essa estabilidade pode ser atribuída à personalização dos modelos para cada adolescente.

Por outro lado, embora o modelo MT1DCNN tenha apresentado um desempenho inicialmente inferior, mostrou uma melhora gradual ao longo do tempo, alcançando taxas de Acurácia Balanceada mais elevadas, à medida que mais dados históricos foram adicionados ao treinamento.

A Figura 32 apresenta o desempenho dos modelos nomotéticos e idiográficos para prever a intensidade dos sintomas da depressão em diferentes agregações considerando as taxas de erro MAE e RMSE.

Os experimentos realizados na tarefa de regressão indicam um desempenho estável dos modelos ao longo do tempo e elevadas taxas de erro MAE e RMSE. Contudo, as taxas de erro diminuíram à medida que mais dados históricos foram incorporados ao aprendizado dos modelos preditivos.

O modelo MT1DCNN-E alcançou os melhores resultados na tarefa de regressão, superando consistentemente os experimentos nomotéticos. Os resultados indicam que a abordagem idiográfica, que considera as especificidades de cada adolescente, revelou-se mais eficaz na previsão das variações das pontuações do sMFQ, que podem refletir as flutuações dos sintomas da depressão dos adolescentes ao longo do período de coleta de dados.

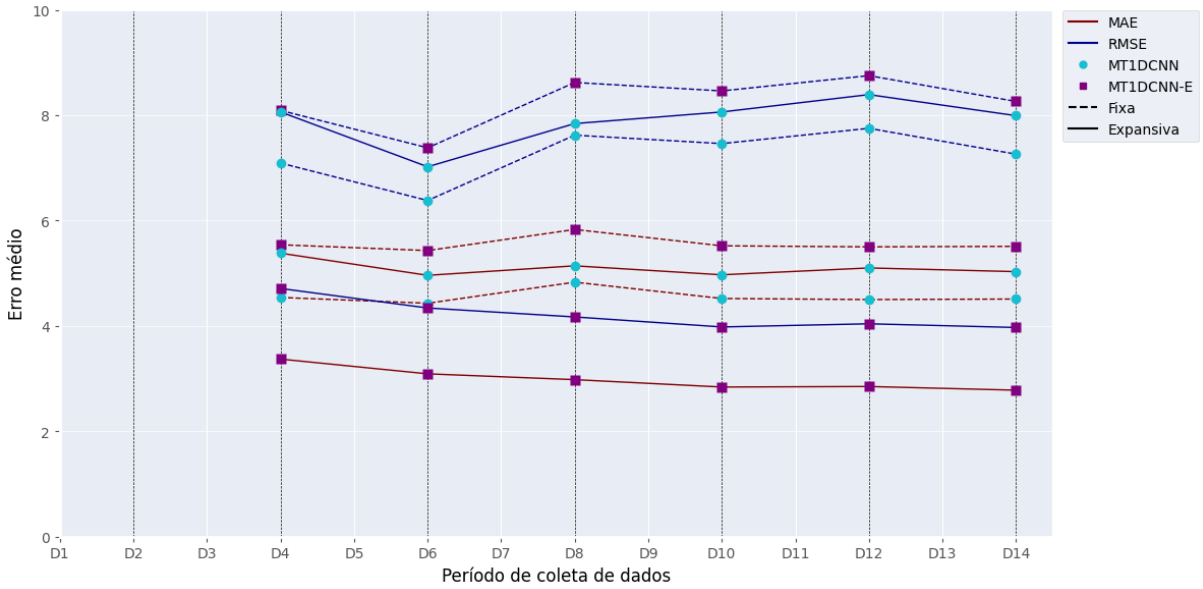


Figura 32 – Desempenho dos modelos nomotético (MT1DCNN) e idiográfico (MT1DCNN-E) na tarefa de regressão para prever os próximos sMFQ. A linha vertical indica o dia em que os sMFQ foram coletados. Os dados correspondentes ao primeiro sMFQ são usados exclusivamente no treinamento do modelo. As taxas de erro são obtidas na previsão dos questionários sMFQ em cada dobra dos conjuntos de validação.

Além disso, as medidas obtidas em ambos os experimentos melhoraram significativamente usando uma abordagem de agregação de dados cumulativa. A Tabela 15 apresenta um resumo das medidas obtidas pelos modelos que se destacaram nas tarefas de classificação binária e de regressão.

Tabela 15 – Melhores experimentos nomotéticos e idiográficos usando agregação cumulativa para a previsão dos próximos sMFQ.

sMFQ	Nomotético			Idiográfico		
	B. Acc	MAE	RMSE	B. Acc	MAE	RMSE
2º	50,00	5,38	8,06	78,39	3,37	4,71
3º	82,41	4,96	7,02	76,90	3,09	4,34
4º	80,39	5,14	7,84	72,32	2,98	4,17
5º	89,81	4,97	8,06	77,40	2,84	3,98
6º	81,48	5,10	8,39	77,98	2,85	4,04
7º	83,11	5,03	7,99	78,15	2,78	3,97

Nossos resultados apontam que a agregação cumulativa contribuiu significativamente para otimizar o desempenho dos modelos preditivos em ambas as tarefas. Além disso, nossas observações sugerem que aumentar o número de pontos de observação ao longo do tempo pode potencialmente aprimorar ainda mais o desempenho dos modelos preditivos.

Contudo, é importante considerar algumas limitações inerentes aos modelos preditivos de depressão desenvolvidos neste estudo. Os modelos nomotéticos são par-

ticularmente úteis quando se pretende identificar padrões comuns em conjunto de dados limitados. Entretanto, é importante notar que a baixa variabilidade dos escores binários dos sMFQ pode ter influenciado positivamente a eficácia dos modelos nomotéticos. Essa baixa variabilidade pode ser atribuída ao curto período de coleta de dados da IDEA-RiSCo de apenas 14 dias, período equivalente à observação de um único episódio depressivo.

Por outro lado, hipotetizamos que os modelos idiográficos demonstraram ser mais eficazes na previsão da intensidade dos sintomas de depressão, ao captarem com mais eficiência as variações individuais no estado de humor dos adolescentes. No entanto,é importante ressaltar que esses modelos demandam um número maior de pontos de observação por indivíduo durante o treinamento, e são difíceis de generalizar devido ao tamanho limitado dos nossos conjuntos de dados.

6 CONCLUSÃO

Neste estudo, exploramos a integração de dados multimodais coletados via *smartphones* e algoritmos de AM para complementar o diagnóstico da depressão em adolescentes brasileiros do IDEA-RiSCo.

Realizamos uma análise comparativa do desempenho dos modelos em tarefas de classificação binária e regressão, usando como base a pontuação do sMFQ e duas agregações de dados com o foco na previsão do humor deprimido atual dos adolescentes. Quatro modalidades de dados foram investigadas individualmente e combinadas visando aprimorar o desempenho dos modelos preditivos. Além disso, nós avaliamos a capacidade de modelos de AM nomotéticos e idiográficos em duas atividades de previsão do humor deprimido futuro.

Nós demonstramos que a integração de diversos tipos de dados coletados por meio de *smartphones* proporciona uma visão mais precisa do estado de humor dos adolescentes. Mostramos que um classificador GB alcançou taxas de Acurácia Balanceada de 86,22% na identificação de amostras deprimidas e não deprimidas ao prever o humor deprimido atual. Além disso, o modelo nomotético (MT1DCNN) alcançou taxas de Acurácia Balanceada mais elevadas, enquanto o modelo idiográfico (MT1DCNN-E) demonstrou maior precisão ao prever os sintomas da depressão, em ambas as atividades de previsão do humor deprimido futuro.

Esses resultados foram obtidos combinando dados de Atividade, Áudios de relatos e Mobilidade. Além disso, a mudança de agregação de dados de uma janela de tempo fixa para uma cumulativa introduziu melhorias significativas na precisão dos classificadores. Portanto, estabelecemos a agregação cumulativa como uma importante alternativa à imputação de dados para minimizar o problema de dados ausentes inerentes aos métodos de coleta passiva.

Além disso, propomos uma discussão sobre quais atributos podem contribuir mais fortemente como marcadores digitais de depressão na adolescência. Dentre os atributos analisados, o tempo de caminhada foi o principal marcador digital de depressão derivado dos dados de atividade. A diminuição do tempo médio de caminhada diária foi associada às amostras deprimidas dos adolescentes em ambas as atividades de

previsão de depressão.

Nós destacamos a relevância dos atributos espectrais, especialmente, MFCC, espectrogramas e contraste como marcadores digitais da fala deprimida. Conduto, os atributos variaram consideravelmente entre as atividades, com uma predominância do MFCC e espectrogramas na previsão do humor deprimido atual. Além disso, observamos que os coeficientes MFCC apresentaram valores superiores nas amostras deprimidas em comparação com as amostras não deprimidas, em ambas as atividades propostas na tese.

Adicionalmente, os atributos derivados da mobilidade que se destacaram foram a distância máxima de casa para a previsão do humor deprimido atual; e a variação de localização na previsão do humor deprimido futuro dos adolescentes.

Identificar os primeiros sinais de depressão na adolescência representa um grande desafio para a sociedade. Nossa pesquisa proporcionou uma visão mais abrangente sobre a contribuição de modalidades de dados para prever o estado de humor dos adolescentes. Além disso, nos permitiu identificar as configurações mais promissoras e os modelos de AM mais eficazes para prever a depressão.

Assim, estabelecemos um modelo multimodal nomotético de AM que integra dados coletados através dos *smartphones* em ambiente doméstico como uma abordagem promissora para enriquecer o diagnóstico da depressão em adolescentes brasileiros.

Considerando que grandes empresas de tecnologias já estão coletando dados de saúde, bem-estar social e *journaling*, acreditamos que modelos multimodais baseados em AM que usam dados coletados via *smartphones* desempenharão um papel cada vez mais importante no suporte a pesquisas voltadas para desafios de saúde mental.

Adicionalmente, nossas descobertas podem ter aplicações práticas no mundo real, abrindo caminho para o desenvolvimento futuro de ferramentas automatizadas de triagem ou acompanhamento que podem auxiliar pais e profissionais de saúde sobre o risco dos adolescentes desenvolverem a depressão nas fases iniciais.

6.1 Contribuições

As principais contribuições deste estudo são:

1. Uma revisão dos principais conjuntos de dados de depressão disponíveis na literatura;

2. Uma análise metodológica de sistemas multimodais baseados em AM para detectar a depressão;

3. Uma pesquisa sobre marcadores digitais de depressão extraídos de *smartphones* e sensores vestíveis;

4. Uma avaliação das estratégias de extração de atributos baseadas em informações acústicas, de atividade e mobilidade.

5. Uma comparação de modelos unimodais e multimodais baseados em aprendizado de máquina para prever a depressão.

6. Uma comparação entre modelos preditivos de depressão usando abordagens nomotéticas e idiográficas.

7. Uma análise dos atributos acústicos (espectrais e prosódicos), atividade e mobilidade como marcadores digitais de depressão.

6.2 Limitações

Nós realizamos um estudo de carácter exploratório baseado em experimentos extensivos para identificar a depressão em adolescentes usando uma série de atributos e quatro modalidades de dados coletados via *smartphones*. Os modelos preditivos desenvolvidos neste estudo possuem alta capacidade preditiva, porém as interpretações dos resultados devem considerar algumas limitações.

O número reduzido de amostras disponíveis, associado aos conjuntos de dados severamente desequilibrados, estão entre os maiores desafios para o desenvolvimento de estudos que aplicam modelos de AM ao contexto da depressão.

Na maioria dos conjuntos de dados, como o DAIC, o Erisk, e AVID-Corpus, observamos que o número de amostras da classe não deprimida é bastante superior ao número de amostras da classe deprimida. Essa distribuição desequilibrada dos conjuntos de dados pode potencialmente introduzir vieses durante o treinamento dos classificadores, um desafio semelhante que encontramos neste estudo. Nossos conjuntos de dados são severamente desequilibrados com uma escassez de amostras da classe deprimida (cerca de 15%) e predominantemente do sexo feminino.

Houve uma grande variabilidade na coleta de dados proveniente dos aplicativos de coleta passiva utilizados neste estudo. Isso resultou na exclusão de alguns adolescentes que tiveram períodos de coletas irregulares em alguns experimentos.

Metodologicamente, adotamos práticas consolidadas na literatura recente para minimizar problemas de sobreajuste durante o treinamento dos modelos preditivos. Utilizamos vários métodos de validação cruzada, otimizadores de hiperparâmetros, regularização, *dropout*, entre outros recursos adotados na maioria dos trabalhos de referência.

Nossos achados de pesquisa são baseados em nossa população e podem refletir um estado momentâneo da saúde mental dos adolescentes. Contudo, o número de amostras reduzido e as próprias características episódicas da depressão podem, possivelmente, limitar a capacidade de generalização dos modelos preditivos.

Cabe ressaltar que as associações entre os atributos e a previsão da classe deprimida, percebidas com base nos gráficos de importância de *Shapley* não implicam necessariamente relações causais no mundo real. Portanto, tais associações podem e devem ser validadas em novas pesquisas de depressão na adolescência.

6.3 Trabalhos Futuros

Como trabalhos futuros, cogitamos explorar outras modalidades de dados coletadas pelos aplicativos IDEA-Bot, GLH e GFIT que estão disponíveis no IDEA-RiSCo. Além disso, consideramos a inserção de uma nova modalidade de dados sociodemográficos dos adolescentes, obtidas nas fases iniciais do estudo IDEA-RiSCo devido à importante contribuição de tais características discutidas em estudos anteriores sobre a depressão (TAZAWA et al., 2020; ASARE et al., 2021; ZHONG et al., 2022; CHIKERSAL et al., 2021; ASARE et al., 2022).

Nós realizamos um estudo sobre a depressão na adolescência com dados coletados no segundo ano do projeto IDEA-RiSCo (W2), portanto, nossos modelos preditivos podem ser replicados para outras ondas. Além disso, planejamos ampliar esta pesquisa, avaliando a capacidade dos modelos para prever o episódio de depressão, através dos questionários MFQ ou K-SADS-PL respondidos ao final do período de coleta de dados do W2 e W3, respectivamente.

Nós cogitamos avaliar estratégias de imputação de dados ausentes para ampliar o número de amostras utilizadas em estudos futuros. Há também oportunidades de introduzir outras arquiteturas de aprendizado profundo, incluindo algumas mais modernas, como redes recorrentes ou modelos híbridos.

Além disso, novas estratégias de extração de atributos serão adotadas em estudos futuros e podem impactar o desempenho dos modelos preditivos. Desse modo, novos atributos espectrais, episódicos, de atividade e mobilidade podem ser investigados como marcadores digitais da depressão na adolescência.

REFERÊNCIAS

AALBERS, G.; HENDRICKSON, A. T.; VANDEN ABEELE, M. M.; KEIJSERS, L. Smartphone-Tracked Digital Markers of Momentary Subjective Stress in College Students: Idiographic Machine Learning Analysis. **JMIR mHealth and uHealth**, [S.l.], v.11, p.e37469, 2023.

AKÇAY, M. B.; OĞUZ, K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. **Speech Communication**, [S.l.], v.116, 2020.

AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY & DATA MINING, 25., 2019. **Proceedings...** [S.l.: s.n.], 2019. p.2623–2631.

ALBUQUERQUE, L. et al. Association between acoustic speech features and non-severe levels of anxiety and depression symptoms across lifespan. **PloS one**, [S.l.], v.16, n.4, p.e0248842, 2021.

ALGHIFARI, M. F. et al. On the Optimum Speech Segment Length for Depression Detection. In: IEEE INTERNATIONAL CONFERENCE ON SMART INSTRUMENTATION, MEASUREMENT AND APPLICATION (ICSIMA), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p.1–5.

ALGHOWINEM, S. From joyous to clinically depressed: Mood detection using multimodal analysis of a person's appearance and speech. In: HUMAINE ASSOCIATION CONFERENCE ON AFFECTIVE COMPUTING AND INTELLIGENT INTERACTION, 2013., 2013. **Anais...** [S.l.: s.n.], 2013. p.648–654.

ALGHOWINEM, S. M. et al. Interpretation of Depression Detection Models via Feature Selection Methods. **IEEE Transactions on Affective Computing**, [S.l.], 2020.

ALMAGHRABI, S. A. **Biomedical Signal Processing For Bioacoustic Features Extraction and Reproducibility Evaluation**. 2022. Tese (Doutorado em Ciência da Computação) — .

ALMEIDA, A. M. G. d. Elaboração de filtros Wavelets baseada em conhecimento para distinçao de locutores autênticos e inautenticos. , [S.l.], 2022.

ANIS, K.; ZAKIA, H.; MOHAMED, D.; JEFFREY, C. Detecting depression severity by interpretable representations of motion dynamics. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE & GESTURE RECOGNITION (FG 2018), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p.739–745.

APA. **Diagnostic and statistical manual of mental disorders (DSM-5®)**. 5th ed..ed. [S.l.]: American Psychiatric Pub, 2013.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. , [S.l.], 2010.

ASARE, K. O. et al. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. **JMIR mHealth and uHealth**, [S.l.], v.9, n.7, p.e26540, 2021.

ASARE, K. O. et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. **Pervasive and Mobile Computing**, [S.l.], p.101621, 2022.

BAILEY, A.; PLUMBLEY, M. D. Raw Audio for Depression Detection Can Be More Robust Against Gender Imbalance than Mel-Spectrogram Features. **arXiv preprint arXiv:2010.15120**, [S.l.], 2020.

BECK, A. T.; STEER, R. A.; BROWN, G. Beck depression inventory–II. **Psychological Assessment**, [S.l.], 1996.

BEHESHTI, I. et al. Histogram-based feature extraction from individual gray matter similarity-matrix for Alzheimer's disease classification. **Journal of Alzheimer's disease**, [S.l.], v.55, n.4, p.1571–1582, 2017.

BELTZ, A. M.; WRIGHT, A. G.; SPRAGUE, B. N.; MOLENAAR, P. C. Bridging the nomothetic and idiographic approaches to the analysis of clinical data. **Assessment**, [S.l.], v.23, n.4, p.447–458, 2016.

BENROUBA, F.; BOUDOUR, R. Emotional sentiment analysis of social media content for mental health safety. **Social Network Analysis and Mining**, [S.l.], v.13, n.1, p.17, 2023.

BERGSTRA, J.; BARDENET, R.; BENGIO, Y.; KÉGL, B. Algorithms for hyper-parameter optimization. **Advances in neural information processing systems**, [S.l.], v.24, 2011.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of machine learning research**, [S.l.], v.13, n.2, 2012.

BERNSTEIN, D. P.; FINK, L.; HANDELSMAN, L.; FOOTE, J. Childhood trauma questionnaire. **Assessment of family violence: A handbook for researchers and practitioners.**, [S.l.], 1998.

BERRAR, D. et al. **Cross-Validation.**

BILAL, A. M. et al. Predicting perinatal health outcomes using smartphone-based digital phenotyping and machine learning in a prospective Swedish cohort (Mom2B): study protocol. **BMJ open**, [S.l.], v.12, n.4, p.e059033, 2022.

BREEBAART, J.; MCKINNEY, M. F. Features for audio classification. **Algorithms in Ambient Intelligence**, [S.l.], p.113–129, 2004.

BREIMAN, L. Random forests. **Machine learning**, [S.l.], v.45, n.1, p.5–32, 2001.

BYUN, S. et al. Entropy analysis of heart rate variability and its application to recognize major depressive disorder: A pilot study. **Technology and Health Care**, [S.l.], v.27, n.S1, p.407–424, 2019.

CAI, H. et al. MODMA dataset: a Multi-model Open Dataset for Mental-disorder Analysis. **arXiv preprint arXiv:2002.09283**, [S.l.], 2020.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, [S.l.], v.16, p.321–357, 2002.

CHEN, X.; SYKORA, M. D.; JACKSON, T. W.; ELAYAN, S. What about mood swings: identifying depression on twitter with temporal measures of emotions. In: COMPANION PROCEEDINGS OF THE THE WEB CONFERENCE 2018, 2018. **Anais...** [S.l.: s.n.], 2018. p.1653–1660.

CHEVANCE, A. et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. **The Lancet Psychiatry**, [S.l.], v.7, n.8, p.692–702, 2020.

CHIKERSAL, P. et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. **ACM Transactions on Computer-Human Interaction (TOCHI)**, [S.l.], v.28, n.1, p.1–41, 2021.

CHOUDHARY, S. et al. A Machine Learning Approach for Detecting Digital Behavioral Patterns of Depression Using Nonintrusive Smartphone Data (Complementary Path to Patient Health Questionnaire-9 Assessment): Prospective Observational Study. **JMIR Formative Research**, [S.l.], v.6, n.5, p.e37736, 2022.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, [S.l.], v.20, n.3, p.273–297, 1995.

COX, J. L.; HOLDEN, J. M.; SAGOVSKY, R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. **The British journal of psychiatry**, [S.l.], v.150, n.6, p.782–786, 1987.

CUMMINS, N. et al. A review of depression and suicide risk assessment using speech analysis. **Speech Communication**, [S.l.], v.71, p.10–49, 2015.

DAVEY, C. G.; MCGORRY, P. D. Early intervention for depression in young people: a blind spot in mental health care. **The Lancet Psychiatry**, [S.l.], v.6, n.3, p.267–272, 2019.

DE CHOUDHURY, M.; COUNTS, S.; HORVITZ, E. Social media as a measurement tool of depression in populations. In: ANNUAL ACM WEB SCIENCE CONFERENCE, 5., 2013. **Proceedings...** [S.l.: s.n.], 2013. p.47–56.

DEPRESSION, W. Other common mental disorders: global health estimates. **Geneva: World Health Organization**, [S.l.], v.24, 2017.

DICK, B.; FERGUSON, B. J. Health for the world's adolescents: a second chance in the second decade. **Journal of Adolescent Health**, [S.l.], v.56, n.1, p.3–6, 2015.

DINKEL, H.; ZHANG, P.; WU, M.; YU, K. Depa: Self-supervised audio embedding for depression detection. **arXiv preprint arXiv:1910.13028**, [S.l.], 2019.

DORYAB, A. et al. Detection of behavior change in people with depression. In: WORKSHOPS AT THE TWENTY-EIGHTH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2014. **Anais...** [S.l.: s.n.], 2014.

DUBAGUNTA, S. P.; VLASENKO, B.; DOSS, M. M. Learning voice source related information for depression detection. In: ICASSP 2019-2019 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2019. **Anais...** [S.l.: s.n.], 2019. p.6525–6529.

ERGUZEL, T. T.; TAS, C.; CEBI, M. A wrapper-based approach for feature selection and classification of major depressive disorder–bipolar disorders. **Computers in biology and medicine**, [S.l.], v.64, p.127–137, 2015.

ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: SECOND INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1996. **Proceedings...** AAAI Press, 1996. p.226–231. (KDD'96).

FERREIRA, D.; KOSTAKOS, V.; DEY, A. K. AWARE: mobile context instrumentation framework. **Frontiers in ICT**, [S.l.], v.2, p.6, 2015.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **Annals of statistics**, [S.l.], v.29, n.5, p.1189–1232, 2001.

GAO, S.; CALHOUN, V. D.; SUI, J. Machine learning in major depression: From classification to treatment outcome prediction. **CNS neuroscience & therapeutics**, [S.l.], v.24, n.11, p.1037–1052, 2018.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. **Machine learning**, [S.l.], v.63, p.3–42, 2006.

GIBBONS, R. D.; DEGRUY, F. V. Without Wasting a Word: Extreme Improvements in Efficiency and Accuracy Using Computerized Adaptive Testing for Mental Health Disorders (CAT-MH). **Current psychiatry reports**, [S.l.], v.21, n.8, p.67, 2019.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.

GRAHAM, A. K. et al. Validation of the computerized adaptive test for mental health in primary care. **The Annals of Family Medicine**, [S.l.], v.17, n.1, p.23–30, 2019.

GRATCH, J. et al. The distress analysis interview corpus of human and computer interviews. In: LREC, 2014. **Anais...** [S.l.: s.n.], 2014. p.3123–3128.

GUNTUKU, S. C.; PREOTIUC-PIETRO, D.; EICHSTAEDT, J. C.; UNGAR, L. H. What Twitter profile and posted images reveal about depression and anxiety. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA, 2019. **Proceedings...** [S.l.: s.n.], 2019. v.13, n.01, p.236–246.

HAMILTON, M. A rating scale for depression. **Journal of neurology, neurosurgery, and psychiatry**, [S.l.], v.23, n.1, p.56, 1960.

HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING, 2005. **Anais...** [S.l.: s.n.], 2005. p.878–887.

HARTE, C.; SANDLER, M.; GASSER, M. Detecting harmonic change in musical audio. In: ACM WORKSHOP ON AUDIO AND MUSIC COMPUTING MULTIMEDIA, 1., 2006. **Proceedings...** [S.l.: s.n.], 2006. p.21–26.

HONG, J. et al. Depressive Symptoms Feature-Based Machine Learning Approach to Predicting Depression Using Smartphone. In: HEALTHCARE, 2022. **Anais...** [S.l.: s.n.], 2022. v.10, n.7, p.1189.

JACOBSON, N. C.; CHUNG, Y. J. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. **Sensors**, [S.l.], v.20, n.12, p.3572, 2020.

JADHAV, N.; SUGANDHI, R. Survey on Human Behavior Recognition Using Affective Computing. In: IEEE GLOBAL CONFERENCE ON WIRELESS COMPUTING AND NETWORKING (GCWCN), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p.98–103.

JIANG, H. et al. Investigation of different speech types and emotions for detecting depression using different classifiers. **Speech Communication**, [S.l.], v.90, p.39–46, 2017.

KANTER, J. W.; BUSCH, A. M.; WEEKS, C. E.; LANDES, S. J. The nature of clinical depression: Symptoms, syndromes, and behavior analysis. **The Behavior Analyst**, [S.l.], v.31, n.1, p.1–21, 2008.

KARYOTAKI, E. et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. **Clinical psychology review**, [S.l.], v.63, p.80–92, 2018.

KIELING, C. et al. Identifying depression early in adolescence. **The Lancet Child & Adolescent Health**, [S.l.], v.3, n.4, p.211–213, 2019.

KIELING, C. et al. The Identifying Depression Early in Adolescence Risk Stratified Cohort (IDEA-RiSCo): rationale, methods, and baseline characteristics. **Frontiers in Psychiatry**, [S.l.], 2021.

KIM, A. Y. et al. Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach. **Journal of Medical Internet Research**, [S.l.], v.25, p.e34474, 2023.

KOVÁCS, G. smote-variants: a Python Implementation of 85 Minority Oversampling Techniques. **Neurocomputing**, [S.l.], v.366, p.352–354, 2019. (IF-2019=4.07).

KOVACS, M. Children's depression inventory. **Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry**, [S.l.], 1992.

KROENKE, K. et al. The PHQ-8 as a measure of current depression in the general population. **Journal of affective disorders**, [S.l.], v.114, n.1-3, p.163–173, 2009.

KUNCHEVA, L. I. A theoretical study on six classifier fusion strategies. **IEEE Transactions on pattern analysis and machine intelligence**, [S.l.], v.24, n.2, p.281–286, 2002.

LAM, G.; DONGYAN, H.; LIN, W. Context-aware Deep Learning for Multi-modal Depression Detection. In: ICASSP 2019-2019 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2019. **Anais...** [S.l.: s.n.], 2019. p.3946–3950.

LAM, R. W. **Depression**. Oxford, UK: Oxford University Press, 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, [S.l.], v.521, n.7553, p.436–444, 2015.

LEE, S. et al. Screening major depressive disorder using vocal acoustic features in the elderly by sex. **Journal of Affective Disorders**, [S.l.], v.291, p.15–23, 2021.

LEMAITRE, G.; NOGUEIRA, F.; ARIDAS, C. K. **Imbalanced-learn**: A python toolbox to tackle the curse of imbalanced datasets in machine learning. 1–5p. v.18, n.17.

LIAW, A.; WIENER, M. et al. Classification and regression by randomForest. **R news**, [S.l.], v.2, n.3, p.18–22, 2002.

LIN, L.; CHEN, X.; SHEN, Y.; ZHANG, L. Towards Automatic Depression Detection: A BiLSTM/1D CNN-Based Model. **Applied Sciences**, [S.l.], v.10, n.23, p.8701, 2020.

LIU, W. et al. A survey of deep neural network architectures and their applications. **Neurocomputing**, [S.l.], v.234, p.11–26, 2017.

LIU, Z.; WANG, D.; ZHANG, L.; HU, B. A Novel Decision Tree for Depression Recognition in Speech. **arXiv preprint arXiv:2002.12759**, [S.l.], 2020.

LOSADA, D. E.; CRESTANI, F. A test collection for research on depression and language use. In: INTERNATIONAL CONFERENCE OF THE CROSS-LANGUAGE EVALUATION FORUM FOR EUROPEAN LANGUAGES, 2016. **Anais...** [S.l.: s.n.], 2016. p.28–39.

LU, L.; ZHANG, H.-J.; JIANG, H. Content analysis for audio classification and segmentation. **IEEE Transactions on speech and audio processing**, [S.l.], v.10, n.7, p.504–516, 2002.

LUNDBERG, S. M.; LEE, S.-I. A Unified Approach to Interpreting Model Predictions. In: GUYON, I. et al. (Ed.). **Advances in Neural Information Processing Systems 30**. [S.l.]: Curran Associates, Inc., 2017. p.4765–4774.

MAGALHAES, T. N.; BARROS, F. B.; LOUREIRO, M. A. Iracema: a python library for audio content analysis. **Revista de Informática Teórica e Aplicada**, [S.l.], v.27, n.4, p.127–138, 2020.

MAHARJAN, S. M. et al. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. **BMC medical informatics and decision making**, [S.l.], v.21, n.1, p.1–19, 2021.

MALHI, G. S.; MANN, J. J. Depression. **The Lancet**, [S.l.], v.392, n.10161, p.2299–2312, 2018.

MASUD, M. T. et al. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. **Journal of biomedical informatics**, [S.l.], v.103, p.103371, 2020.

MASUD, M. T. et al. Non-Pervasive Monitoring of Daily-Life Behavior to Access Depressive Symptom Severity Via Smartphone Technology. In: IEEE REGION 10 SYMPOSIUM (TENSYMP), 2020., 2020. **Anais...** [S.l.: s.n.], 2020. p.602–607.

MAYES, T. L. et al. Psychometric properties of the Children's Depression Rating Scale–Revised in adolescents. **Journal of child and adolescent psychopharmacology**, [S.l.], v.20, n.6, p.513–516, 2010.

MCFEE, B. et al. librosa: Audio and music signal analysis in python. In: OF THE 14TH PYTHON IN SCIENCE CONFERENCE, 2015. **Proceedings...** [S.l.: s.n.], 2015. v.8, p.18–25.

MEGHANANI, A.; ANOOP, C.; RAMAKRISHNAN, A. An exploration of log-mel spectrogram and MFCC features for Alzheimer's dementia recognition from spontaneous speech. In: IEEE SPOKEN LANGUAGE TECHNOLOGY WORKSHOP (SLT), 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p.670–677.

MIN, S. et al. Acoustic analysis of speech for screening for suicide risk: machine learning classifiers for between-and within-person evaluation of suicidality. **Journal of medical internet research**, [S.l.], v.25, p.e45456, 2023.

MITCHELL, A. J.; VAZE, A.; RAO, S. Clinical diagnosis of depression in primary care: a meta-analysis. **The Lancet**, [S.l.], v.374, n.9690, p.609–619, 2009.

MONTGOMERY, S. A.; ÅSBERG, M. A new depression scale designed to be sensitive to change. **The British journal of psychiatry**, [S.l.], v.134, n.4, p.382–389, 1979.

MOSHE, I. et al. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. **Frontiers in psychiatry**, [S.l.], v.12, 2021.

MULLICK, T. et al. Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning–Based Exploratory Study. **JMIR Formative Research**, [S.l.], v.6, n.6, p.e35807, 2022.

MUNDT, J. C. et al. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. **Journal of neurolinguistics**, [S.l.], v.20, n.1, p.50–64, 2007.

NATARAJAN, S. et al. Boosting for postpartum depression prediction. In: IEEE/ACM INTERNATIONAL CONFERENCE ON CONNECTED HEALTH: APPLICATIONS, SYSTEMS AND ENGINEERING TECHNOLOGIES (CHASE), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p.232–240.

NGUYEN, B.; KOLAPPAN, S.; BHAT, V.; KRISHNAN, S. Clustering and feature analysis of smartphone data for depression monitoring. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE & BIOLOGY SOCIETY (EMBC), 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p.113–116.

NGUYEN, D.-K.; CHAN, C.-L.; ADAMS LI, A.-H.; PHAN, D.-V. Deep Stacked Generalization Ensemble Learning models in early diagnosis of Depression illness from wearable devices data. In: INTERNATIONAL CONFERENCE ON MEDICAL AND HEALTH INFORMATICS, 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p.7–12.

PAMPOUCHIDOU, A. et al. Quantitative comparison of motion history image variants for video-based depression assessment. **EURASIP Journal on Image and Video Processing**, [S.l.], v.2017, n.1, p.64, 2017.

PAPPALARDO, L. et al. Returners and explorers dichotomy in human mobility. **Nature communications**, [S.l.], v.6, n.1, p.1–8, 2015.

PAPPALARDO, L. et al. An analytical framework to nowcast well-being using mobile phone data. **International Journal of Data Science and Analytics**, [S.l.], v.2, n.1, p.75–92, 2016.

PAPPALARDO, L.; SIMINI, F.; BARLACCHI, G.; PELLUNGRINI, R. **scikit-mobility**: a Python library for the analysis, generation and risk assessment of mobility data.

PATEL, M. J.; KHALAF, A.; AIZENSTEIN, H. J. Studying depression using imaging and machine learning methods. **NeuroImage: Clinical**, [S.l.], v.10, p.115–123, 2016.

PATIAS, N. D.; MACHADO, W. D. L.; BANDEIRA, D. R.; DELL'AGLIO, D. D. Depression Anxiety and Stress Scale (DASS-21)-short form: adaptação e validação para adolescentes brasileiros. **Psico-USF**, [S.l.], v.21, p.459–469, 2016.

PAUL, T. et al. **Modeling Human Mobility Entropy as a Function of Spatial and Temporal Quantizations**. 2017. Tese (Doutorado em Ciência da Computação) — University of Saskatchewan.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, [S.l.], v.12, p.2825–2830, 2011.

PEDRELLI, P. et al. Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors. **Frontiers in psychiatry**, [S.l.], v.11, p.1413, 2020.

POSNER, K. et al. Columbia-suicide severity rating scale (C-SSRS). **New York, NY: Columbia University Medical Center**, [S.l.], v.10, p.2008, 2008.

POZNANSKI, E. O.; COOK, S. C.; CARROLL, B. J. A depression rating scale for children. **Pediatrics**, [S.l.], v.64, n.4, p.442–450, 1979.

POZNANSKI, E. O.; MOKROS, H. Children's depression rating scale, revised (CDRS-R). **Western Psychological Services**, [S.l.], v.18, n.2, p.213–217, 1996.

PROKHORENKOVA, L. et al. CatBoost: unbiased boosting with categorical features. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2018. **Anais...** [S.l.: s.n.], 2018. p.6638–6648.

RADLOFF, L. S. The CES-D scale: A self-report depression scale for research in the general population. **Applied psychological measurement**, [S.l.], v.1, n.3, p.385–401, 1977.

REECE, A. G.; DANFORTH, C. M. Instagram photos reveal predictive markers of depression. **EPJ Data Science**, [S.l.], v.6, n.1, p.1–12, 2017.

REJAIBI, E. et al. MFCC-based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech. **arXiv preprint arXiv:1909.07208**, [S.l.], 2019.

RENDIMENTO, B. M. da Economia. Instituto Brasileiro de Geografia e Estatísticas (IBGE). Diretoria de Pesquisas Coordenação de Trabalho e. **Pesquisa Nacional de Saúde 2019**: Percepção do estado de saúde, estilos de vida e doenças crônicas e saúde bucal-Brasil e grandes regiões. [S.l.]: IBGE Rio de Janeiro, 2020.

RENDIMENTO, I. B. de Geografia e Estatística. Coordenação de Trabalho e. **Pesquisa Nacional de Saúde 2013**: percepção do estado de saúde, estilos de vida e doenças crônicas. [S.l.]: Ministério de Planejamento, Orçamento e Gestão, Instituto Brasileiro de . . . , 2014.

RINGEVAL, F. et al. Avec 2017: Real-life depression, and affect recognition workshop and challenge. In: ANNUAL WORKSHOP ON AUDIO/VISUAL EMOTION CHALLENGE, 7., 2017. **Proceedings. . .** [S.l.: s.n.], 2017. p.3–9.

ROCHA, T. B.-M. et al. Identifying adolescents at risk for depression: A prediction score performance in cohorts based in 3 different continents. **Journal of the American Academy of Child & Adolescent Psychiatry**, [S.l.], v.60, n.2, p.262–273, 2021.

ROSA, M.; METCALF, E.; ROCHA, T. B.-M.; KIELING, C. Translation and cross-cultural adaptation into Brazilian portuguese of the mood and feelings questionnaire (MFQ)–long version. **Trends in psychiatry and psychotherapy**, [S.l.], v.40, n.1, p.72–78, 2018.

RUSSELL, J. A. A circumplex model of affect. **Journal of personality and social psychology**, [S.l.], v.39, n.6, p.1161, 1980.

RYKOV, Y. et al. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. **JMIR mHealth and uHealth**, [S.l.], v.9, n.10, p.e24872, 2021.

SAEB, S. et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. **Journal of medical Internet research**, [S.l.], v.17, n.7, p.e175, 2015.

SAEB, S. et al. The relationship between mobile phone location sensor data and depressive symptom severity. **PeerJ**, [S.l.], v.4, p.e2537, 2016.

SALVATORE, S.; VALSINER, J. Between the general and the unique: Overcoming the nomothetic versus idiographic opposition. **Theory & Psychology**, [S.l.], v.20, n.6, p.817–833, 2010.

SCHWARTZ, H. A. et al. Towards assessing changes in degree of depression through facebook. In: OF THE WORKSHOP ON COMPUTATIONAL LINGUISTICS AND CLINICAL PSYCHOLOGY: FROM LINGUISTIC SIGNAL TO CLINICAL REALITY, 2014. **Proceedings. . .** [S.l.: s.n.], 2014. p.118–125.

SHAH, R. V. et al. Personalized machine learning of depressed mood using wearables. **Translational Psychiatry**, [S.l.], v.11, n.1, p.1–18, 2021.

SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, [S.l.], v.27, n.3, p.379–423, 1948.

SHARMA, G.; UMAPATHY, K.; KRISHNAN, S. Trends in audio signal feature extraction methods. **Applied Acoustics**, [S.l.], v.158, p.107020, 2020.

SHARMA, P.; RAJPOOT, A. K. Automatic identification of silence, unvoiced and voiced chunks in speech. **Journal of Computer Science & Information Technology (CS & IT)**, [S.l.], v.3, n.5, p.87–96, 2013.

SHUAI, H.-H. et al. A comprehensive study on social network mental disorders detection via online social media mining. **IEEE Transactions on Knowledge and Data Engineering**, [S.l.], v.30, n.7, p.1212–1225, 2017.

SMITH, L.; FOLEY, L.; PANTER, J. Activity spaces in studies of the environment and physical activity: A review and synthesis of implications for causality. **Health & place**, [S.l.], v.58, p.102113, 2019.

SNAITH, R. P. et al. A scale for the assessment of hedonic tone the Snaith–Hamilton Pleasure Scale. **The British Journal of Psychiatry**, [S.l.], v.167, n.1, p.99–103, 1995.

SONG, C.; QU, Z.; BLUMM, N.; BARABÁSI, A.-L. Limits of predictability in human mobility. **Science**, [S.l.], v.327, n.5968, p.1018–1021, 2010.

STASAK, B.; EPPS, J.; CUMMINS, N.; GOECKE, R. An Investigation of Emotional Speech in Depression Classification. In: INTERSPEECH, 2016. **Anais...** [S.l.: s.n.], 2016. p.485–489.

STEPANOV, E. A. et al. Depression severity estimation from multiple modalities. In: IEEE 20TH INTERNATIONAL CONFERENCE ON E-HEALTH NETWORKING, APPLICATIONS AND SERVICES (HEALTHCOM), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p.1–6.

STOPA, S. R. et al. Pesquisa Nacional de Saúde 2019: histórico, métodos e perspectivas. **Epidemiologia e Serviços de Saúde**, [S.l.], v.29, 2020.

STRINGARIS, A. et al. The Affective Reactivity Index: a concise irritability scale for clinical and research settings. **Journal of Child Psychology and Psychiatry**, [S.l.], v.53, n.11, p.1109–1117, 2012.

SUMALI, B. et al. Speech quality feature analysis for classification of depression and dementia patients. **Sensors**, [S.l.], v.20, n.12, p.3599, 2020.

SUWANNAKHUN, S.; YINGTHAWORNSUK, T. Characterizing Depressive Related Speech with MFCC. In: INTERNATIONAL JOINT SYMPOSIUM ON ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING (ISAI-NLP), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p.1–6.

TAGUCHI, T. et al. Major depressive disorder discrimination using vocal acoustic features. **Journal of affective disorders**, [S.l.], v.225, p.214–220, 2018.

TAZAWA, Y. et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. **Heliyon**, [S.l.], v.6, n.2, p.e03274, 2020.

THAPAR, A.; EYRE, O.; PATEL, V.; BRENT, D. Depression in young people. **The Lancet**, [S.l.], 2022.

TØNNING, M. L. et al. Mood and activity measured using smartphones in unipolar depressive disorder. **Frontiers in Psychiatry**, [S.l.], p.1135, 2021.

VALSTAR, M. et al. AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. In: ACM INTERNATIONAL WORKSHOP ON AUDIO/VISUAL EMOTION CHALLENGE, 3., 2013. **Proceedings...** [S.l.: s.n.], 2013. p.3–10.

VÁSQUEZ-CORREA, J. C.; OROZCO-ARROYAVE, J.; BOCKLET, T.; NÖTH, E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. **Journal of communication disorders**, [S.l.], v.76, p.21–36, 2018.

VELAZQUEZ, B. et al. Physical activity and depressive symptoms among adolescents in a school-based sample. **Brazilian Journal of Psychiatry**, [S.l.], v.44, p.313–316, 2022.

VIDUANI, A. et al. Assessing Mood With the Identifying Depression Early in Adolescence Chatbot (IDEABot): Development and Implementation Study. **JMIR human factors**, [S.l.], v.10, p.e44388, 2023.

WANG, R. et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: ACM INTERNATIONAL JOINT CONFERENCE ON PERVASIVE AND UBIQUITOUS COMPUTING, 2014., 2014. **Proceedings...** [S.l.: s.n.], 2014. p.3–14.

WANG, W. et al. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, [S.l.], v.2, n.3, p.1–21, 2018.

WANG, Y. et al. Fast and accurate assessment of depression based on voice acoustic features: a cross-sectional and longitudinal study. **Frontiers in Psychiatry**, [S.l.], v.14, p.1195276, 2023.

WANG, Y.; YALCIN, A.; VANDEWEERD, C. An entropy-based approach to the study of human mobility and behavior in private homes. **PLoS one**, [S.l.], v.15, n.12, p.e0243503, 2020.

WARE, S. et al. Automatic depression screening using social interaction data on smartphones. **Smart Health**, [S.l.], v.26, p.100356, 2022.

WILLIAMSON, J. R. et al. Tracking depression severity from audio and video based on speech articulatory coordination. **Computer Speech & Language**, [S.l.], v.55, p.40–56, 2019.

YAN, Y.; TU, M.; WEN, H. A CNN Model with Discretized Mobile Features for Depression Detection. In: IEEE-EMBS INTERNATIONAL CONFERENCE ON WEARABLE AND IMPLANTABLE BODY SENSOR NETWORKS (BSN), 2022., 2022. **Anais...** [S.l.: s.n.], 2022. p.1–4.

YANG, L. et al. Hybrid depression classification and estimation from audio video and text information. In: ANNUAL WORKSHOP ON AUDIO/VISUAL EMOTION CHALLENGE, 7., 2017. **Proceedings...** [S.l.: s.n.], 2017. p.45–51.

YANG, Y.; FAIRBAIRN, C.; COHN, J. F. Detecting depression severity from vocal prosody. **IEEE transactions on affective computing**, [S.l.], v.4, n.2, p.142–150, 2012.

YATES, A.; COHAN, A.; GOHARIAN, N. Depression and self-harm risk assessment in online forums. **arXiv preprint arXiv:1709.01848**, [S.l.], 2017.

YINGTHAWORNSUK, T. Spectral entropy in speech for classification of depressed speakers. In: INTERNATIONAL CONFERENCE ON SIGNAL-IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS (SITIS), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p.679–682.

YUSOF, N. F. A.; LIN, C.; GUERIN, F. Analysing the causes of depressed mood from depression vulnerable individuals. In: INTERNATIONAL WORKSHOP ON DIGITAL DISEASE DETECTION USING SOCIAL MEDIA 2017 (DDDSM-2017), 2017. **Proceedings...** [S.l.: s.n.], 2017. p.9–17.

ZHAO, Q. et al. Vocal Acoustic Features as Potential Biomarkers for Identifying/Diagnosing Depression: A Cross-Sectional Study. **Frontiers in Psychiatry**, [S.l.], v.13, p.815678, 2022.

ZHONG, M. et al. Unimodal vs. Multimodal Prediction of Antenatal Depression from Smartphone-based Survey Data in a Longitudinal Study. In: INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION, 2022., 2022. **Proceedings...** [S.l.: s.n.], 2022. p.455–467.

ZHOU, S. et al. GPS2space: An Open-source Python Library for Spatial Measure Extraction from GPS Data. **Journal of Behavioral Data Science**, [S.l.], v.1, n.2, p.127–155, 2021.

ZHOU, X.; JIN, K.; SHANG, Y.; GUO, G. Visually interpretable representation learning for depression recognition from facial images. **IEEE Transactions on Affective Computing**, [S.l.], 2018.

ZUNG, W. W.; RICHARDS, C. B.; SHORT, M. J. Self-rating depression scale in an outpatient clinic: further validation of the SDS. **Archives of general psychiatry**, [S.l.], v.13, n.6, p.508–515, 1965.

Apêndices

APÊNDICE A – Análise metodológica dos trabalhos relacionados

A.0.1 *Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone*

O trabalho de Masud et al. (2020), investigou padrões de atividade e movimento como marcadores digitais de depressão. Eles coletaram dados de acelerômetro e GPS de 33 participantes adultos (19 homens e 14 mulheres) durante 11 semanas por meio do aplicativo *Data Collector*, instalado nos *smartphones* Android dos participantes.

Eles avaliaram um total de 12 atributos, de atividade: em repouso; pisar; "caminhar"; se exercitar; tempo de transição; uso do telefone; uso do telefone na posição deitada e ficar em casa. De localização: variação de distância; movimento circadiano; entropia e entropia normalizada. Mais detalhes podem ser encontrados em Masud et al. (2020).

Os experimentos compararam os algoritmos *k-means*, Máquina de Vetores de Suporte (SVM - do inglês *Support Vector Machine*) e uma rede neural artificial do tipo perceptron de multicamadas (MLP - do inglês *multi-layer perceptron*) em uma tarefa de classificação de três níveis do PHQ-9: Ausência de depressão (pontuação ≤ 10), depressão moderada (pontuação entre 10 e 15) e depressão grave (pontuação ≥ 15).

Os modelos foram treinados para prever os escores semanais de depressão do PHQ-9. O modelo com melhor resultado utilizou o classificador SVM, obteve uma Acurácia de 87,2% para prever indivíduos com sintomas graves de depressão.

Além disso, os autores utilizaram um método de seleção de atributos (wrapper), adotado na pesquisa de Erguzel; Tas; Cebi (2015), para avaliar quais subconjuntos de atributos foram mais relevantes para o aprendizado do modelo. Eles concluíram que níveis mais graves de depressão estavam mais fortemente correlacionados com menor variação de distância, mais tempo de descanso, mais tempo em casa, uso mais mentiroso do telefone, menos passos, menos entropia normalizada e menos exercício.

A.0.2 *Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning*

Tazawa et al. (2020) desenvolveram um modelo de AM para rastrear e identificar a depressão usando dados de 86 indivíduos adultos com idade média de 59 anos (45 pacientes com depressão e 41 pacientes saudáveis) selecionados em diferentes

clínicas médicas no Japão. Profissionais especializados avaliaram previamente os participantes usando HAMD, MADRS e BDI, entre outros questionários psiquiátricos.

Os autores monitoraram os idosos diariamente para medir o número de passos, gasto de energia, movimento corporal, tempo de sono, frequência cardíaca, temperatura da pele e exposição à luz ultravioleta usando uma pulseira *Silmee W11*. Eles coletaram características sociodemográficas como idade, sexo, duração da doença, diagnósticos e medicamentos durante a avaliação clínica.

Os pesquisadores coletaram dados vestíveis sete dias antes da avaliação clínica. A frequência da coleta varia conforme o tipo de dado. No entanto, eles agregaram os atributos por hora para treinar o modelo preditivo. Ao final, o estudo avaliou 63 atributos extraídos de sensores vestíveis. Mais detalhes sobre o processo de coleta e extração de atributos podem ser encontrados em Tazawa et al. (2020).

Os autores usaram dois conjuntos de dados. Um conjunto de dados contendo dados coletados em três dias, outro conjunto de dados contendo dados coletados em sete dias. Além de três modelos de AM, SVM, XGB, *Random Forest*. Os melhores resultados foram alcançados utilizando XGBoost.

Para o conjunto de dados de 3 dias, os autores obtiveram uma Acurácia de 74%, Sensibilidade de 66% e Especificidade de 81%. Para o conjunto de dados de 7 dias, alcançaram uma Acurácia de 76%, Sensibilidade de 73% e Especificidade de 79%.

Os resultados indicam que a temperatura da pele e o tempo de sono foram as características mais fortes para o aprendizado do modelo. Os autores também sugerem diferenças significativas na frequência cardíaca, passos e sinais de sono de participantes deprimidos e saudáveis.

A.0.3 Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors

Pedrelli et al. (2020) usaram dados obtidos via sensores vestíveis e *smartphones* de 31 pacientes com idades entre 19 e 71 anos para estimar a gravidade da depressão. Os pacientes completaram seis questionários HDRS durante oito semanas e foram previamente diagnosticados com depressão.

Sensores vestíveis: Os participantes usaram duas pulseiras *Empathic E4*, uma em cada braço, para medir a atividade eletrotérmica, temperatura da pele, medições de domínio de tempo e frequência de variabilidade da frequência cardíaca, tempo em movimento e características do sono (hora de início do sono, tempo máximo ininterrupto de sono, duração do sono, número de despertares noturnos, horário de despertar, índice de regularidade do sono).

Sensores do smartphone: Os participantes instalaram o aplicativo *MovisensXS* que foi responsável pela coleta de dados de localização, ligações telefônicas, SMS, Tela, dados meteorológicos e aplicativos utilizados durante o período de monitora-

mento.

O estudo avaliou três modelos de AM para prever a gravidade dos sintomas de depressão do HDRS: o primeiro combinava atributos de *smartphones* e pulseiras, o segundo usava apenas atributos de *smartphones* e o terceiro usava apenas atributos de pulseiras. Os autores agruparam os dados em intervalos de seis horas e diários.

O modelo que utilizou apenas atributos extraídos dos *smartphones* dos participantes teve melhor desempenho em ambos os casos. Foi avaliado pelo erro médio absoluto, obtendo valores entre 3,88 e 4,74.

A.0.4 *Predicting Depression From Smartphone Behavioral Markers Using Machine Learning Methods, Hyperparameter Optimization, and Feature Importance Analysis: Exploratory Study*

O estudo proposto por Asare et al. (2021) explorou a relação entre 22 atributos calculados com base no Índice de regularidade (WANG et al., 2018); entropia (SHANNON, 1948); desvio padrão e contagens. Os atributos foram extraídos dos dados de *status* da tela, da conectividade com a internet e dos logs de uso de aplicativos em primeiro plano usando o aplicativo *Carat* instalado nos *smartphones Android* dos participantes.

Os autores calcularam os atributos com base na agregação diária dos dados dos participantes. O conjunto de dados consiste em 13.898 dias de coleta para 629 participantes de 56 países, com uma média de 22,1 dias de coleta por participante e 1.600 avaliações do PHQ-8. O conjunto de dados foi validado com base na nota de corte do PHQ-8. Uma pontuação do PHQ-8 $\geq$ 10 para a classe deprimida e uma pontuação do PHQ-8 <10 para classe não deprimida.

Os autores avaliaram cinco algoritmos de AM: floresta aleatória, SVM, XGB, KNN e regressão logística em uma tarefa de classificação binária de depressão. Eles adotaram técnicas de otimização de hiperparâmetros, validação cruzada estratificada e aninhada de 10 vezes e sobreamostragem de amostras da classe minoritária.

Eles usaram dois conjuntos de dados. Um formado apenas com atributos extraídos de *smartphones* e outro com a adição de variáveis sociodemográficas, idade e sexo dos participantes.

Os resultados indicam que a inclusão de características sociodemográficas no conjunto de dados melhorou o desempenho dos modelos desenvolvidos. O modelo com o melhor desempenho utilizou o classificador XGBoost e alcançou taxas de precisão de 94,03%, Revocação, 95,62%; F1, 95,27%; área sob a curva 99,36%; e Acurácia de 94,93% para prever a classe deprimida.

O estudo utilizou um método de importância de atributos para identificar quais marcadores foram mais significativos para o aprendizado dos modelos preditivos. Os atributos de conectividade de tela e conectividade com a internet foram mais fortes para

prever a depressão. Além disso, os resultados apontaram uma associação positiva significativa entre a entropia normalizada pelo *status* da tela e a depressão.

A.0.5 *Detecting Depression and Predicting its Onset Using Longitudinal Symptoms Captured by Passive Sensing: A Machine Learning Approach With Robust Feature Selection*

Chikersal et al. (2021) propõe um modelo de AM multitarefa para prever depressão em estudantes universitários após o final do semestre letivo (ausência ou presença) e avaliar a mudança na intensidade dos sintomas (piora ou não) a partir de dados de longo prazo coletados durante o semestre.

Os autores usaram o aplicativo AWARE (FERREIRA; KOSTAKOS; DEY, 2015) instalado nos *smartphones Android* para coletar passivamente dados de 138 alunos por 16 semanas (um semestre). Eles extraíram uma série de atributos com base em cinco categorias de dados: *Bluetooth*, chamadas telefônicas, localização, mapa do campus e uso do telefone. Eles também usaram rastreadores de fitness da pulseira *Fitbit Flex* 2 para extrair dados sobre passos e sono. Além disso, os alunos responderam a dois questionários BDI-II na primeira e na última semana do semestre, que serviram como verdade básica para validar os modelos preditivos.

Metodologicamente, o modelo proposto utiliza um algoritmo de seleção de atributos e os algoritmos regressão logística e *Gradient Boosting*. Os autores adotaram o método de validação cruzada de exclusão (deixar uma pessoa de fora) para otimizar os hiperparâmetros durante o treinamento do modelo preditivo com base nas tarefas de classificação de depressão pós-semestre e de alteração da depressão no decorrer do semestre. Com isso, o sistema pôde explorar, além de métricas de desempenho, os atributos mais frequentemente escolhidos pelo modelo de seleção.

O modelo que obteve o melhor desempenho permite detectar a presença de depressão pós-semestre com uma Acurácia de 85,7% usando os conjuntos de atributos: *Bluetooth*, Chamadas telefônica, Uso do telefone e Passos e a mudança de depressão com uma Acurácia de 85,4% usando os conjuntos de atributos: *Bluetooth*, Mapa do campus, Uso do telefone e Sono. Além disso, os resultados apontam que os modelos conseguiram prever esses resultados com uma Acurácia > 80%, 11 a 15 semanas antes do final do semestre.

A.0.6 *Clustering and Feature Analysis of Smartphone Data for Depression Monitoring*

Nguyen et al. (2021) propuseram um modelo de AM para monitorar os níveis de gravidade da depressão. Os autores agruparam os dados dos participantes usando um algoritmo *bi-clustering multi-view* para prever três níveis de pontuações do PHQ-9 (baixo, médio e alto).

Os autores usaram dados de 48 estudantes universitários do *Student Life*. Eles calcularam os atributos com base na agregação diária de dados. São eles:

- **Atividade**: Atividades parado, caminhada e corrida.
- **Conversação**: Número de conversas e tempo total de conversas realizadas pelo *smartphone*
- **Luz**: O tempo total e número de vezes em que um usuário está em um ambiente escuro.
- **Áudio**: Duração total dos segmentos de áudio dos usuários, classificados como silêncio, ruído e voz.
- **Bloqueio do telefone**: Número de vezes que o telefone é bloqueado e a duração total do bloqueio.

Os autores utilizaram uma técnica de seleção de atributos baseada no critério de redundância mínima e relevância máxima para reduzir o conjunto de atributos que melhor distinguem as classes do PHQ-9. Como resultado, o conjunto de dados total tinha 33 atributos, enquanto o conjunto de dados otimizado foi reduzido para os 16 atributos mais importantes.

O modelo preditivo formado por um classificador *Decision tree* treinado com o conjunto de dados reduzido de 16 atributos e validação cruzada de dez vezes, atingiu uma Acurácia de 94,7% para classificar os 3 grupos de pontuação do PHQ-9 em estudantes universitários do *StudentLife*.

Os resultados indicam que os estudantes universitários com baixa pontuação PHQ-9 foram associados a mais atividades de caminhadas e conversas realizadas pelo *smartphone*.

A.0.7 *Depressive Symptoms Feature-Based Machine Learning Approach to Predicting Depression Using Smartphone*

Hong et al. (2022) propuseram um modelo multimodal para prever a depressão em pacientes coreanos recrutados em clínicas psiquiátricas. Eles coletaram dados por duas semanas usando o aplicativo *Mental Health Protector* instalado nos *smartphones* dos participantes. O conjunto de dados contém 33 atributos extraídos de dados de sono, atividade, e expressão facial de 106 participantes.

- **Sono**: Foram calculados 14 atributos para estimar as características do sono ativando e desativando a tela do celular. Eles usaram média, desvio padrão, máximo, mínimo estimado de tempo de sono por dia (quantidade de sono) e uso de *smartphone* por dia (qualidade do sono).

- **Atividade**: São formados por atributos de mobilidade (variação de localização, entropia) e do tempo médio diário para as atividades, caminhada, corrida, parado.

- **Expressão facial**: Os autores utilizaram uma rede *EfficientNet* pré-treinada com o conjunto de dados *Facial Expression Comparison* (FEC) para classificar as imagens faciais dos participantes, fotografados semanalmente em emoções como raiva, concentração, nojo e tristeza, entre outros.

Os autores desenvolveram um modelo preditor de depressão baseado no algoritmo de AM *Random Forest* para prever pontuações binárias de depressão do PHQ-9 e CESD-R. Eles demonstraram que o uso de atributos de expressão facial permitiu alcançar valores de precisão mais altos dos escores binários do PHQ-9 (66,67%) do que atributos de sono (55,56%) e atividade física (59,26%), isoladamente. No entanto, a combinação dos atributos foi crucial para melhorar o desempenho do classificador, atingindo 74,07% e 76,92% de Acurácia nas pontuações binárias PHQ-9 e CESD-R, respectivamente.

A.0.8 *A Machine Learning Approach for Detecting Digital Behavioral Patterns of Depression Using Nonintrusive Smartphone Data (Complementary Path to Patient Health Questionnaire-9 Assessment): Prospective Observational Study*

Choudhary et al. (2022) propôs uma métrica de perfil comportamental mental (MHSS - do inglês - *Mental Health Similarity Score*) para prever a ausência e a presença de depressão e os níveis de gravidade dos sintomas usando as pontuações do PHQ-9.

O conjunto de dados foi coletado usando o *Behavior Research App* instalado nos *smartphones* Android de 558 participantes sul-coreanos. O estudo utilizou 37 atributos extraídos com base nas 11 categorias do *behavidence*. Os atributos adotados no trabalho estão descritas abaixo:

- Tempo gasto pelos usuários nos aplicativos de cada categoria do *behavidence* (App 0 a App 11)

- Tempo em que os participantes interagem com o celular (*Mean session, total session*)

- Tempo em que os participantes não estão interagindo com o celular (*sleep, average gap*)

- Número de vezes em que os participantes abrem os aplicativos de cada categoria (*App 1- Number of opens a app10 - Number of opens*),

- Número de vezes que os participantes abrem os aplicativos de cada categoria e teve tempos de sessão maiores que a média da sessão da categoria (*App 1- Upper limits a App 10 - Upper Limits*)

- Tempo médio de movimento dos participantes com base no giroscópio (*Average Activity, Average Gap Activity, Total Activity*).

Eles agregaram os dados coletados a cada 24 horas. Mais detalhes sobre coleta de dados e atributos podem ser acessados em Choudhary et al. (2022).

Os autores realizaram vários experimentos usando os algoritmos de AM *multivariate adaptive regression splines, random forest*, XGB, e SVM. Além disso, eles avaliaram quatro modelos preditivos usando o questionário PHQ-9 como base: um modelo binário padrão (nenhum ou grave), outro modelo binário adicionando atributos de giroscópio, um modelo de 3 classes (nenhum, moderado ou grave) e um modelo binário treinado nas questões 2, 6 e 9.

Os melhores resultados são obtidos pelo modelo de 2 classes usando o classificador *Random Forest*, que atingiu uma taxa de Acurácia de 87%. E pelo modelo de 3 classes (nenhum, moderada ou grave) que alcançou uma Acurácia de 78%.

Os resultados apontam que o tempo médio da sessão e o uso de aplicativos de interação social (categoria 1) foram maiores em participantes com depressão grave em comparação aos participantes sem depressão.

A.0.9 *Unimodal vs Multimodal Prediction of Antenatal Depression from Smartphone-based Survey Data in a Longitudinal Study*

Zhong et al. (2022) desenvolveram um estudo sobre a identificação precoce de depressão em gestantes de língua sueca. Os autores utilizaram uma coorte de 915 mulheres maiores de 18 anos, com dados coletados no primeiro e segundo trimestres de gravidez através do aplicativo Mob2b (BILAL et al., 2022) e completaram o EPDS entre a 36ª e 42ª semanas de gestação.

O conjunto de dados contém dados sociodemográficos, saúde psicológica, saúde geral, comportamento, social e personalidade. Além disso, os autores realizaram experimentos unimodais e multimodais em uma tarefa de classificação binária de depressão pré-natal.

Os autores adotaram o método de busca em grade com validação cruzada estratificada de 5 vezes nos experimentos unimodais e compararam o desempenho de cinco algoritmos de AM: SVM, *Logistic Regression* (LR), *K-Nearest Neighbor* (KNN), XGBoost e MLP.

Além disso, aplicaram duas operações de fusão sobre os atributos para construir os vetores de entradas dos modelos multimodais: uma fusão de nível de atributo que concatena os atributos das modalidades de dados em único vetor de entrada e uma

fusão de nível de decisão no qual as modalidades de dados são treinadas separadamente, em seguida são aplicadas cinco estratégias de fusão de algoritmos: fixa, média, máxima e mediana e de aprendizado supervisionado (regressão logística, MLP, Gaussiana). Mais detalhes sobre as estratégias de fusão podem ser encontrados em Zhong et al. (2022); Kuncheva (2002)

Os autores usaram um método de importância para destacar os atributos que otimizam o desempenho dos modelos unimodais e multimodais. Dados de saúde psicológica contribuíram mais para detectar depressão precoce em mulheres grávidas. Além disso, os modelos multimodais superaram os modelos unimodais, em geral.

O modelo que utilizou regra de fusão de decisão com regressão logística obteve os melhores resultados, atingindo uma Acurácia Balanceada de 75% e uma área sob a curva de 82%. Ao final, a prevalência de depressão foi de 21% na amostra pesquisada.

A.0.10 *Predicting Depression in Adolescents Using Mobile and Wearable Sensors: Multimodal Machine Learning-Based Exploratory Study*

Mullick et al. (2022) avaliaram estratégias de modelagem de dados universais e personalizadas para prever escores de depressão e a intensidade dos sintomas de depressão. Um grupo de 37 adolescentes com idade entre 12 e 17 anos, com média de idade de 15,5 anos, participaram do estudo. Seus dados foram coletados passivamente durante 24 semanas por meio de sensores de *smartphones* e dispositivos vestíveis.

O estudo avaliou estratégias de modelagem do conjunto de dados baseadas em validação cruzada de exclusão durante o treinamento dos modelos de AM. Especificamente, eles adotaram duas estratégias universais: Deixar um participante de fora; Deixar semana X de fora; e duas estratégias personalizadas: Deixar uma instância de usuário de uma semana; e Semanas acumuladas. Mais detalhes sobre as estratégias universais e personalizadas podem ser consultadas em (MULLICK et al., 2022).

Os autores usaram o aplicativo *AWARE* para coletar o uso da tela do telefone, chamadas, conversas, Wi-Fi e dados de localização a cada 10 minutos. Eles também usaram o *Fitbit Inspire HR* (versão 1.84.5) para coletar dados de frequência cardíaca, sono e passos com uma frequência de uma vez a cada minuto. Eles avaliaram 66 atributos e sua correlação com a sintomatologia da depressão. Além disso, os adolescentes completaram o PHQ-9 semanalmente para validar o conjunto de dados de depressão.

A modelagem que obteve o melhor desempenho, Semanas acumuladas associada a um classificador XGBoost, pôde prever os escores de depressão e a variação semanal do nível de depressão com valores de MAE de 2,39 e 3,21, respectivamente.

Os resultados indicaram a contribuição de atributos baseados em tela, chamada e localização como melhores preditores da depressão em adolescentes. Além disso,

os autores sugerem que modelos personalizados, tiveram melhor desempenho na previsão da depressão em adolescentes do que as abordagens universais.

A.0.11 *A CNN Model with Discretized Mobile atributos for Depression Detection*

O trabalho de Yan; Tu; Wen (2022) propõe um modelo baseado em aprendizado profundo que usa dados sequenciais coletados via *smartphones* e pulseiras eletrônicas para prever a ausência ou presença de depressão em adultos chineses.

O conjunto de dados contém dados de 263 adultos chineses com idades entre 20 e 60 anos, coletados durante cinco dias. Eles usaram a agregação diária dos dados para calcular dez atributos: uso do telefone, atividade física, dieta, sono e localização. Além disso, os participantes completaram o DASS-21 (PATIAS et al., 2016) que serviu para validar o modelo desenvolvido.

Os autores adotaram uma rede neural convolucional unidimensional (1DCNN) associada a um método de discretização apoiado no Valor da Informação, que agrupa os atributos com base em um valor de confiança da informação.

Esse método de Discretização usa regras para avaliar a importância dos atributos e dividi-los em 4 grupos com base no valor da informação (IV - do inglês Information Value): imprevisível (IV<0,02), fraco (IV: 0,02 a 0,1), médio (IV: 0,1 a 0,3) e forte (IV: 0,3+). Os atributos com maiores IV foram número de vezes que usou o telefone (0,300), tempo de duração do uso do telefone (0,286) e calorias ingeridas (0,181).

Os autores usaram a pesquisa em grade para encontrar o conjunto ideal de hiperparâmetros do modelo. Eles compararam o desempenho de sete modelos de previsão de depressão: SVM, MLP, *Decision Tree, Random Forest e Logistic Regression* e 1DCNN com e sem o uso do método de discretização.

O modelo 1DCNN que utiliza o método de discretização dos atributos com base no IV obteve os melhores resultados, alcançando valores de Acurácia de 83%, Medida-F1 de 79% e AUC de 80%.

A.0.12 *Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis*

Asare et al. (2022) avaliam uma série de atributos extraídos de *smartphones* e sensores vestíveis (anel Oura) como marcadores de depressão. Um total de 54 participantes foram recrutados em comunidades online e monitorados por 30 dias. As avaliações de humor foram coletadas usando o Modelo Circumplex de Afeto (RUSSELL, 1980) que conceitua o humor em uma escala *likert* para dimensões de Valência e Excitação. Os atributos e modalidades avaliados neste estudo estão descritos abaixo:

- **Sono**: Tempo total de sono, movimento rápido dos olhos, latência de início do

sono, acordar após o início do sono, eficiência do sono, variabilidade da frequência cardíaca

- **Atividade Física**: Contagem de passos e equivalente metabólico para tarefa
- **Mobilidade**: Variação de localização, distância total, entropia de localização, entropia de localização normalizada, entropia de localização de log, velocidade média, movimento circadiano, número de lugares significativos, raio de giro, movimento para proporção estática, tempo em casa, e índice de Regularidade de Localização
- **Uso do telefone**: Duração do uso do telefone e a frequência de uso do telefone, índice de Regularidade da tela e o ritmo Circadiano da Tela
- **Humor**: Valência negativa e positiva; Excitação negativa e positiva

Os autores avaliaram o desempenho de cinco algoritmos de AM: SVM, *Random forest*, XGBoost, KNN e Regressão Logística com método de validação cruzada aninhada em uma tarefa de classificação binária de depressão.

O modelo que obteve o melhor desempenho usou dados sociodemográficos, *smartphones* e sensores vestíveis e avaliações de humor. Ele conseguiu discriminar classes deprimidas e não deprimidas com valores de Acurácia (81,43%), AUC (82,31%) e F1 macro (73,34%).

Além disso, adotaram o método *SHapley Additive exPlanations* (SHAP) para selecionar os 45 atributos que mais contribuíram para otimizar o desempenho do classificador. Relacionados ao uso do telefone: soma, média, desvio padrão da duração do desbloqueio da tela e contagem de desbloqueios da tela. Relacionados a mobilidade: tempo médio de permanência em locais significativos, número de locais significativos, índice de rotina de localização e localização normalizada entropia. Para a modalidade do Sono: Tempo Total de Sono e Eficiência do Sono.

A.0.13 *Automatic depression screening using social interaction data on smartphones*

Ware et al. (2022) propuseram um sistema de triagem automática que utiliza dados de interação social para prever a presença ou ausência de depressão. O estudo utilizou 59 estudantes universitários com idade entre 18 e 25 anos. Os dados de interação social são baseados em padrões de uso de mensagens SMS e registros de chamadas telefônicas capturados dos *smartphones* dos participantes ao longo de 14 dias.

O conjunto de dados de SMS é formado de 29 atributos entre contagem de mensagens, contatos únicos das mensagens, comprimento das mensagens, entropia e entropia normalizada das mensagens e do comprimento das mensagens, média e

desvio padrão do comprimento das mensagens. Os pesquisadores calcularam essas medidas para mensagens recebidas e enviadas; e para os períodos da manhã, tarde e noite. O conjunto de dados de chamadas telefônicas contém 30 atributos, incluindo informações sobre a quantidade, duração, entropias, médias e desvio padrão das chamadas telefônicas recebidas e enviadas da mesma maneira que o conjunto de dados SMS.

Os atributos de interação social foram calculados usando dados agrupados em intervalos diários. A lista completa dos atributos de interação social extraídos de SMS e chamadas telefônicas pode ser consultada com mais detalhes em Ware et al. (2022).

Os autores desenvolveram três experimentos para prever a pontuação de depressão do HDRS: o primeiro usou apenas atributos de SMS; o segundo usou somente atributos de chamada telefônica e o terceiro usou atributos dos dois tipos de dados de interação social.

Eles adotaram um método de validação cruzada de exclusão (deixar uma pessoa de fora) no treinamento do modelo preditivo, utilizando os algoritmos de AM, *Random Forest*, SVM e XGBoost.

Os experimentos indicam que a combinação de dados de SMS e chamadas telefônicas melhorou o desempenho dos modelos desenvolvidos. O modelo de melhor resultado utilizou o classificador XGboost e obteve taxa de Acurácia de 80% na tarefa de classificação binária de depressão em estudantes universitários.

APÊNDICE B – Atributos

Os processos de engenharia de atributos descritos no Capítulo 4 foram importantes para reduzir a complexidade e a dimensionalidade dos dados utilizados na pesquisa. Adotamos as bibliotecas *Librosa* e Iracema para extrair atributos espectrais. Usamos o *Disvoice* para extrair atributos prosódicos e as bibliotecas SKMOB e GPS2Space para extrair os dados de mobilidade. Os dados de atividades são interpretados pelo EBM por esse motivo não foi necessário utilizar ferramentas especialistas.

A Tabela 16 apresenta a descrição dos atributos espectrais adotados neste trabalho. Estes atributos foram selecionados a partir de um conjunto maior, após uma fase de testes preliminares, em que foram analisados transformadas de *fourier*, coeficiente delta (diferencial) e delta-delta (aceleração), número de coeficientes cepstrais de MFCC e Espectrogramas, entre outros parâmetros configuração. Alguns atributos precisam de mais de coeficientes para representar um sinal de áudio.

As Tabelas 17, 18 e 19 apresentam a descrição dos atributos prosódicos, de atividade e mobilidade adotados neste trabalho, respectivamente. Neste estudo, investigamos 81 atributos extraídos a partir de dados de smartphones como marcadores digitais da depressão na adolescência.

Tabela 16 – Atributos espectrais extraídos dos áudios de relatos. O * indica que o atributo é formado por mais de um coeficiente espectral.

Nome	Descrição	Unidade de Medida	Valores permitidos
w2_bot_spe_zero_crossing_rate	Taxa de cruzamentos por zero de um sinal por segundo. Um ZCR é mais alta em sinais ruidosos do que em harmônicos (MCFEE et al., 2015).	Numérico	≥ 0
w2_bot_spe_rolloff	Calcula a frequência em que um sinal fica abaixo de 85% da energia espectral total, indicando sons audíveis e não audíveis. Um *roll-off* alto indica tom, um baixo indica ruído (ALMAGHRABI, 2022).	Numérico	$-\infty$ - $+\infty$
w2_bot_spe_contrast	Representa a distribuição relativa das frequências do sinal de fala. Ela mede a diferença de amplitude entre picos (conteúdo harmônico) e vales espectrais (componentes de ruído) (LU; ZHANG; JIANG, 2002).	Numérico	$-\infty$ - $+\infty$
w2_bot_spe_tonnals*	Detecta mudanças harmônicas medindo escalas maiores e menores de um sinal de fala que podem ser associadas a sentimentos de felicidade e tristeza, respectivamente (HARTE; SANDLER; GASSER, 2006).	Numérico	$-\infty$ - $+\infty$
w2_bot_spe_bandwidths*	Mede a variação do centróide espectral de um sinal de fala para discriminar sons tonais de sons semelhantes a ruído (ALMEIDA, 2022)	Numérico	$-\infty$ - $+\infty$
w2_bot_spe_mfcc_accelerate*	Coeficientes cepstrais de frequência mel de 40 dimensões usando a derivada de segunda ordem (BREEBAART; MCKINNEY, 2004)	Numérico	$-\infty$ - $+\infty$

w2_bot_spe_log_mel_spectrograms*	Descreve a energia do cepstrum em uma escala não linear (mel) de 128 dimensões escaladas em decibéis (MEGHANANI; ANOOP; RAMAKRISHNAN, 2021)	Numérico	-80 - +80
w2_bot_spe_centroid	Média ponderada das frequências de um sinal de fala. Indica a frequência predominante do sinal (ALMEIDA, 2022).	Numérico	- ∞ - + ∞
w2_bot_spe_spread	Calcula o desvio padrão do espectro em torno do centróide espectral (LU; ZHANG; JIANG, 2002). Mede a propagação do espectro em um sinal de voz	Numérico	- ∞ - + ∞
w2_bot_spe_skewness	Mede a assimetria da distribuição do espectro em torno do centróide espectral. A assimetria aumenta para segmentos sonoros e cai para zero durante as pausas (SHARMA; UMAPATHY; KRISHNAN, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_kurtosis	Mede a distribuição de pico do espectro, valores negativos sugerem uma distribuição mais plana e valores positivos sugerem uma distribuição com um pico mais nítido (SHARMA; UMAPATHY; KRISHNAN, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_entropy	Mede o grau de imprevisibilidade da distribuição espectral de um sinal. Os segmentos de fala têm valores de entropia mais baixos em comparação com segmentos de ruído ou sem fala (ALMAGHRABI, 2022).	Numérico	- ∞ - + ∞
w2_bot_spe_energy	Calcula a energia total de um sinal de voz (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_rms_energy	Calcula a energia média do sinal pela variação de amplitude (MCFEE et al., 2015).	Numérico	- ∞ - + ∞

w2_bot_spe_harmonic_energy	Calcula a energia das parciais harmônicas de um sinal de fala (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_harmonic_centroid	Calcula o centro de gravidade das amplitudes dos sinais harmônicos da fala (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_noisiness	Representa a relação entre o nível de energia do ruído (valores próximos a 1) e a energia total de um sinal (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_hfc	Calcula o conteúdo de alta frequência de um sinal (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_peak	Calcula o valor de amplitude de pico absoluto de um sinal (MAGALHAES; BARROS; LOUREIRO, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_flatness	Mede a estabilidade da fala quantificando partes com voz e sem voz do sinal de áudio. Valores próximos a um indicam ruído e valores próximos a zero indicam um tom (SHARMA; UMAPATHY; KRISHNAN, 2020).	Numérico	- ∞ - + ∞
w2_bot_spe_flux	Mede a velocidade das mudanças na distribuição de energia de quadros sonoros sucessivos. É inversamente proporcional ao nivelamento espectral (ALMEIDA, 2022).	Numérico	- ∞ - + ∞
w2_bot_spe_pitch	Compara componentes harmônicos de um sinal usando o método *Harmonic Product Spectrum* (HPS). Valores altos de *pitch* sugerem um som mais agudo, enquanto valores mais baixos, um som mais grave	Numérico	- ∞ - + ∞

Tabela 17 – Atributos prosódicos extraídos a partir dos áudios episódicos capturados no ambiente

Nome	Descrição	Unidade de Medida	Valores permitidos
w2_ebm_epi_F0avg	Média do contorno F0 dos sinais de áudio	Numérico	≥ 0
w2_ebm_epi_F0std	Desvio padrão do contorno F0 dos sinais de áudio	Numérico	≥ 0
w2_ebm_epi_F0kurt	Curtose do contorno F0 dos sinais de áudio	Numérico	- ∞ - + ∞
w2_ebm_epi_F0mseavg	Valor médio do MSE de F0 para sinais de áudio	Numérico	≥ 0
w2_ebm_epi_F0mseskw	Desvio padrão do MSE da F0 de um sinal de áudio	Numérico	≥ 0
w2_ebm_epi_F0msestd	Assimetria do MSE da F0 dos sinais de áudio	Numérico	- ∞ - + ∞
w2_ebm_epi_F0tiltavg	Inclinação média da F0 dos sinais de áudio	Numérico	- ∞ - + ∞
w2_ebm_epi_F0tiltskw	Assimetria de inclinação da F0 dos sinais de áudio	Numérico	≥ 0
w2_ebm_epi_F0tiltstd	Desvio padrão de inclinação da F0 dos sinais de áudio	Numérico	≥ 0
w2_ebm_epi_avgdurvoiced	Duração média dos sons vocais	Numérico (segundos)	≥ 0
w2_ebm_epi_stddurvoiced	Desvio padrão da duração dos sons vocais	Numérico	≥ 0
w2_ebm_epi_skwdurvoiced	Assimetria da duração dos sons vocais	Numérico	- ∞ - + ∞
w2_ebm_epi_Avgdurunvoiced	Duração média dos sons não vocais	Numérico (segundos)	≥ 0
w2_ebm_epi_stddurunvoiced	Desvio padrão da duração dos sons não vocais	Numérico	≥ 0
w2_ebm_epi_Skwdurunvoiced	Assimetria da duração dos sons não vocais	Numérico	- ∞ - + ∞
w2_ebm_epi_avgdurpause	Duração média da pausa	Numérico (segundos)	≥ 0
w2_ebm_epi_stddurpause	Desvio padrão da duração da pausa	Numérico	≥ 0
w2_ebm_epi_skwdurpause	Assimetria da duração da pausa	Numérico	- ∞ - + ∞

Tabela 18 – Atributos de atividade extraídos a partir dos dados do EBM

Variável	Descrição	Unidade de Medida	Valores permitidos
w2_ebm_act_In_Vehicle	Soma das atividades In_vehicle	Inteiro	≥ 0
w2_ebm_act_On_Bicycle	Soma das atividades On_Bicycle	Inteiro	≥ 0
w2_ebm_act_On_Foot	Soma das atividades On_Foot	Inteiro	≥ 0
w2_ebm_act_Still	Soma das atividades Still	Inteiro	≥ 0
w2_ebm_act_Tilting	Soma das atividades Inclinação	Inteiro	≥ 0
w2_ebm_act_Unknown	Soma de atividades Unknown	Inteiro	≥ 0
w2_ebm_act_unique	Quantidade de atividades distintas realizadas pelo participante	Inteiro	≥ 0
w2_ebm_act_count	Soma de todas as atividades	Inteiro	≥ 0
w2_ebm_act_sumDurIn_Vehicle	Duração total da atividade In_vehicle	Numérico (segundos)	≥ 0
w2_ebm_act_sumDurOn_Bicycle	Duração total da atividade On_Bicycle	Numérico (segundos)	≥ 0
w2_ebm_act_sumDurOn_Foot	Duração total da atividade On_Foot	Numérico (segundos)	≥ 0
w2_ebm_act_sumDurStill	Duração total da atividade Still	Numérico (segundos)	≥ 0
w2_ebm_act_sumDurTilting	Duração total da atividade Tilting	Numérico (segundos)	≥ 0
w2_ebm_act_sumDurUnknown	Duração total da atividade Unknown	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurIn_Vehicle	Duração média da atividade In_vehicle	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurOn_Bicycle	Duração média da atividade On_Bicycle	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurOn_Foot	Duração média da atividade On_Foot	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurStill	Duração média da atividade Still	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurTilting	Duração média da atividade Tilting	Numérico (segundos)	≥ 0
w2_ebm_act_avgDurUnknown	Duração média da atividade Unknown	Numérico (segundos)	≥ 0

Tabela 19 – Atributos de mobilidade extraídos a partir dos dados de localização (GPS)

Nome	Descrição	Unidade de Medida	Valores permitidos
w2_ebm_mob_maximum_distance	Distância máxima percorrida	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_distance_straight_line	Distância máxima percorrida em linha reta	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_max_distance_from _home	Distância máxima percorrida de casa (local de referência entre as 22:00 e 07:00).	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_location_variance	Variação dos pontos de localização (latitude e longitude) do adolescente (SAEB et al., 2015).	Numérico	- ∞ - + ∞
w2_ebm_mob_number_of_visits	Número total visitas	Inteiro	≥ 0
w2_ebm_mob_number_of_locations	Número de locais distintos visitados	Inteiro	≥ 0
w2_ebm_mob_number_of_clusters	Número de locais agrupados (*clusters*).	Inteiro	≥ 0
w2_ebm_mob_random_entropy	Mede o grau de previsibilidade do movimento do adolescente considerando que cada local visitado tem a mesma probabilidade de ser selecionado (SONG et al., 2010).	Numérico	≥ 0
w2_ebm_mob_uncorrelated_entropy	Calculada com base em uma distribuição de probabilidades que considera a frequência em que os locais foram visitados (SONG et al., 2010).	Numérico	≥ 0
w2_ebm_mob_real_entropy	Calculada com base na frequência de visitação, na ordem em que os locais foram visitados e o tempo gasto em cada local (PAPPALARDO et al., 2019).	Numérico	≥ 0

w2_ebm_mob_buff_area	Calcula o espaço de atividade com base na área da circunferência (raio de 100 metros) dos pontos de localização (ZHOU et al., 2021).	Numérico (Metro quadrado)	≥ 0
w2_ebm_mob_convex_area	Calcula o espaço de atividade com base na área do polígono formado pelos pontos de localização mais extremos (ZHOU et al., 2021).	Numérico (Metro quadrado)	≥ 0
w2_ebm_mob_radius_of_gyration	Mede a distância característica percorrida pelos adolescentes, induzida pelos locais visitados com maior frequência (PAPPALARDO et al., 2016).	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_5k_radius_of_gyration	Mede a distância característica percorrida com base nos seus 5 locais mais frequentes (PAPPALARDO et al., 2016)	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_avg_wt	Tempo médio gasto no deslocamento entre os pontos de localização	Numérico (Segundos)	≥ 0
w2_ebm_mob_avg_jl	Distância média de deslocamentos entre os pontos de localização	Numérico (Quilômetro)	≥ 0
w2_ebm_mob_n_trips	Total de caminhadas realizadas	Inteiro	≥ 0
w2_ebm_mob_day_trips	Número de caminhadas realizadas durante o dia	Inteiro	≥ 0
w2_ebm_mob_night_trips	Número de caminhadas realizadas durante a noite	Inteiro	≥ 0
w2_ebm_mob_avg_stop_trips_days	Média de caminhadas diárias.	Inteiro	≥ 0

APÊNDICE C – Experimentos

Os experimentos foram realizados nos conjuntos de dados de treinamento e teste, com foco nas tarefas de classificação binária e regressão. A classificação binária considera a pontuação de corte do sMFQ adotada durante a validação dos dados do IDEA-RiSCo para os conjuntos de dados agregados. A regressão considera a pontuação total do sMFQ que vai de um intervalo entre 0 e 23 em nossos conjuntos de dados.

Além disso, uma série de estratégias foram avaliadas durante o desenvolvimento dos modelos preditivos: técnicas de sobreamostragem de dados foram utilizadas para balanceamento das classes; alterações no tamanho das sequências (quadros sonoros) dos áudios episódicos e de relatos; treinamento de dados normalizados e não normalizados; avaliação dos atributos individualmente e em conjunto; e pesos também atribuídos às funções de perda durante o treinamento para ajustar o modelo.

C.1 Previsão do humor deprimido atual

Nós conduzimos os experimentos em duas configurações: unimodais e multimodais. Nos experimentos unimodais foram utilizadas apenas uma modalidade de dados. Os experimentos multimodais são formados a partir das combinações entre as modalidades.

Os dados foram distribuídos em uma proporção de 75% / 25% para os conjuntos de dados de treinamento e teste com divisão estratificada das amostras, seguindo o estabelecido na metodologia. Adotamos esses experimentos para avaliar a capacidade dos modelos preditivos para prever os escores binários de depressão com base no sMFQ.

As Tabelas 20 e 21 apresentam os resultados dos experimentos de Previsão do humor deprimido atual em ambas as agregações de dados, respectivamente. Eles são apresentados em ordem decrescente, com base no valor de Acurácia Balanceada no conjunto de testes.

C.1.1 Tarefa: Previsão dos escores binários do questionário sMFQ

Tabela 20 – Humor deprimido atual: Previsão dos escores binários do sMFQ usando agregação fixa.

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
ARM	1DCNN	92,84	83,46	84,26	78,95	81,6	58,81
AE	GB	83,93	88,98	92,58	68,42	80,51	65,0
AER	1DCNN	90,16	85,04	87,03	73,68	80,36	59,57
AM	1DCNN	89,8	84,25	87,03	68,42	77,73	56,52
Relatos	DNN	85,19	80,31	81,47	73,68	77,58	52,83
ER	1DCNN	91,60	71,65	69,44	84,21	76,83	47,06
ERM	DNN	85,71	85,83	89,81	63,16	76,49	57,14
AR	1DCNN	92,67	85,04	88,89	63,16	76,02	55,81
AERM	1DCNN	92,10	88,19	93,52	57,89	75,71	59,46
ERM	1DCNN	92,31	84,25	87,96	63,16	75,56	54,55
RM	ET	94,07	88,89	94,95	55,55	75,25	60,61
Relatos	1DCNN	92,24	86,61	91,67	57,89	74,78	56,41
RM	SVM	73,24	88,03	93,94	55,55	74,75	58,81
RM	CB	82,87	83,76	87,88	61,11	74,49	53,66
AEM	GB	84,37	88,98	95,37	52,62	74,0	58,81
ARM	DNN	88,1	84,25	88,89	57,89	73,39	52,38
AERM	SVM	75,39	87,4	93,52	52,62	73,08	55,55
Relatos	SVM	73,83	87,4	93,52	52,62	73,08	55,55
RM	BG	86,38	83,76	88,89	55,55	72,22	51,28
ARM	SVM	74,48	81,89	86,11	57,89	72,0	48,89
RM	1DCNN	90,64	87,18	93,94	50,0	71,97	54,55
ERM	SVM	74,36	81,10	85,19	57,89	71,54	47,83
Atividade	CB	81,74	88,19	95,37	47,37	71,37	54,55
AE	CB	84,0	86,61	93,52	47,37	70,44	51,43
AER	SVM	74,8	86,61	93,52	47,37	70,44	51,43
AM	CB	81,25	86,61	93,52	47,37	70,44	51,43
AM	GB	84,24	86,61	93,52	47,37	70,44	51,43
ER	RF	94,62	86,61	93,52	47,37	70,44	51,43
AERM	CB	87,16	85,83	92,58	47,37	69,98	50,0
AER	DNN	89,11	85,04	91,67	47,37	69,52	48,65
AERM	DNN	88,55	85,04	91,67	47,37	69,52	48,65
AR	SVM	75,29	84,25	90,74	47,37	69,05	47,37
AR	CB	86,72	86,61	94,44	42,11	68,27	48,48
RM	RF	91,84	80,34	85,86	50,0	67,93	43,9

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
Mobilidade	DNN	74,0	63,78	62,03	73,68	67,86	37,84
Atividade	RF	89,51	85,83	93,52	42,11	67,81	47,06
AR	DNN	92,88	85,83	93,52	42,11	67,81	47,06
ER	GB	88,89	81,89	87,96	47,37	67,67	43,9
AERM	GB	91,13	85,04	92,58	42,11	67,35	45,71
ARM	RF	93,99	85,04	92,58	42,11	67,35	45,71
AEM	CB	83,11	87,4	96,3	36,84	66,57	46,67
Atividade	GB	79,64	83,46	90,74	42,11	66,42	43,24
ARM	CB	87,24	83,46	90,74	42,11	66,42	43,24
ER	CB	86,00	83,46	90,74	42,11	66,42	43,24
AM	SVM	67,43	79,53	85,19	47,37	66,28	40,91
Mobilidade	RF	87,6	79,53	85,19	47,37	66,28	40,91
AE	RF	93,24	86,61	95,37	36,84	66,11	45,16
AER	CB	86,86	86,61	95,37	36,84	66,11	45,16
AERM	RF	94,67	86,61	95,37	36,84	66,11	45,16
ER	DNN	89,45	86,61	95,37	36,84	66,11	45,16
Atividade	1DCNN	87,3	82,67	89,81	42,11	65,96	42,11
Relatos	GB	89,05	82,67	89,81	42,11	65,96	42,11
Relatos	RF	93,24	82,67	89,81	42,11	65,96	42,11
Mobilidade	GB	76,89	78,74	84,26	47,37	65,81	40,0
Mobilidade	1DCNN	60,2	52,76	47,22	84,21	65,72	34,78
AER	GB	89,86	85,83	94,44	36,84	65,64	43,75
ARM	GB	89,01	85,83	94,44	36,84	65,64	43,75
Atividade	DNN	80,42	81,89	88,89	42,11	65,5	41,03
AEM	RF	93,17	88,98	99,07	31,58	65,33	46,15
AR	ET	96,89	88,98	99,07	31,58	65,33	46,15
ERM	CB	85,81	85,04	93,52	36,84	65,18	42,42
AERM	ET	96,96	88,19	98,15	31,58	64,86	44,44
AEM	DNN	84,11	84,25	92,58	36,84	64,72	41,18
AEM	1DCNN	84,3	80,31	87,03	42,11	64,57	39,01
RM	DNN	84,61	82,05	89,9	38,89	64,39	40,0
AM	DNN	83,59	83,46	91,67	36,84	64,25	40,0
ARM	ET	96,77	86,61	96,3	31,58	63,94	41,38
Relatos	ET	96,95	86,61	96,3	31,58	63,94	41,38
ERM	GB	87,94	85,83	95,37	31,58	63,47	40,0
AE	SVM	71,67	81,89	89,81	36,84	63,33	37,84

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
AEM	SVM	72,5	81,89	89,81	36,84	63,33	37,84
AR	GB	88,59	85,04	94,44	31,58	63,01	38,71
ER	SVM	74,36	81,10	88,89	36,84	62,87	36,84
AER	ET	96,33	88,19	99,07	26,32	62,69	40,0
EM	DNN	80,01	84,25	93,52	31,58	62,55	37,5
EM	SVM	67,08	75,59	81,47	42,11	61,79	34,04
ER	ET	97,13	85,83	96,3	26,32	61,30	35,70
ERM	RF	93,23	85,83	96,3	26,32	61,30	35,70
Relatos	CB	84,67	81,89	90,74	31,58	61,16	34,29
Atividade	ET	94,02	85,04	95,37	26,32	60,84	34,48
AE	1DCNN	91,3	85,04	95,37	26,32	60,84	34,48
AM	RF	92,97	85,04	95,37	26,32	60,84	34,48
AE	DNN	84,37	81,10	89,81	31,58	60,69	33,33
Mobilidade	ET	91,89	77,17	84,26	36,84	60,55	32,56
AR	RF	94,08	84,25	94,44	26,32	60,38	33,33
Mobilidade	SVM	58,13	69,28	73,15	47,37	60,26	31,58
ERM	ET	95,87	86,61	98,15	21,05	59,59	32,0
AER	RF	94,11	85,83	97,22	21,05	59,14	30,76
EM	GB	79,9	81,10	90,74	26,32	58,53	29,40
EM	ET	96,26	87,4	100,0	15,79	57,89	27,27
AM	ET	96,35	83,46	94,44	21,05	57,75	27,58
Mobilidade	CB	77,55	75,59	83,33	31,58	57,46	27,91
EM	RF	91,39	81,89	92,58	21,05	56,82	25,81
EM	1DCNN	87,9	81,10	91,67	21,05	56,36	25,0
Episódicos	1DCNN	75,1	77,17	86,11	26,32	56,21	25,64
Episódicos	GB	76,5	80,31	90,74	21,05	55,90	24,24
Atividade	SVM	64,74	75,59	84,26	26,32	55,28	24,39
AEM	ET	97,19	85,04	98,15	10,53	54,33	17,39
AE	ET	97,53	84,25	97,22	10,53	53,87	16,66
EM	CB	79,99	81,89	94,44	10,53	52,49	14,81
Episódicos	RF	89,31	81,10	93,52	10,53	52,01	14,29
Episódicos	ET	94,31	82,67	96,3	5,26	50,78	8,33
Episódicos	CB	78,64	78,74	90,74	10,53	50,62	12,9
Episódicos	SVM	64,52	73,22	83,33	15,79	49,55	15,0
Episódicos	DNN	76,42	74,8	86,11	10,53	48,32	11,11

Tabela 21 – Humor deprimido atual: Previsão dos escores binários do sMFQ usando agregação cumulativa.

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
ARM	GB	89,8	92,47	95,16	77,27	86,22	75,56
AR	DNN	87,03	85,61	85,48	86,36	85,92	64,41
RM	GB	89,82	91,78	94,35	77,27	85,81	73,91
AERM	RF	95,19	93,84	97,58	72,72	85,15	78,05
ARM	RF	94,89	93,15	96,77	72,72	84,75	76,19
AERM	CB	88,08	92,47	95,97	72,72	84,35	74,42
AM	CB	85,33	92,47	95,97	72,72	84,35	74,42
AR	RF	94,88	92,47	95,97	72,72	84,35	74,42
AER	CB	89,39	91,78	95,16	72,72	83,94	72,72
ARM	CB	88,5	91,78	95,16	72,72	83,94	72,72
Relatos	RF	94,47	91,78	95,16	72,72	83,94	72,72
RM	CB	88,5	91,78	95,16	72,72	83,94	72,72
AR	CB	88,23	90,41	93,55	72,72	83,14	69,57
ERM	GB	90,55	90,41	93,55	72,72	83,14	69,57
Relatos	CB	87,92	89,73	92,74	72,72	82,73	68,08
AER	RF	94,69	92,47	96,77	68,17	82,48	73,17
AR	ET	96,21	92,47	96,77	68,17	82,48	73,17
Relatos	ET	95,92	92,47	96,77	68,17	82,48	73,17
Relatos	DNN	85,06	89,03	91,94	72,72	82,33	66,67
AER	DNN	85,69	79,45	78,23	86,36	82,28	55,87
ERM	DNN	86,59	79,45	78,23	86,36	82,28	55,87
RM	RF	94,93	91,78	95,97	68,17	82,07	71,43
AE	CB	85,38	91,10	95,16	68,17	81,67	69,77
EM	CB	86,91	91,10	95,16	68,17	81,67	69,77
Episódicos	RF	93,7	91,10	95,16	68,17	81,67	69,77
ER	CB	90,13	91,10	95,16	68,17	81,67	69,77
Mobilidade	CB	78,79	91,10	95,16	68,17	81,67	69,77
AERM	DNN	87,39	90,41	94,35	68,17	81,27	68,17
ERM	RF	94,78	90,41	94,35	68,17	81,27	68,17
ERM	CB	88,53	90,41	94,35	68,17	81,27	68,17
AR	1DCNN	92,86	90,41	94,35	68,17	81,27	68,17
EM	SVM	70,88	86,99	89,52	72,72	81,12	62,74
Episódicos	ET	96,48	92,47	97,58	63,63	80,61	71,78
RM	ET	95,78	92,47	97,58	63,63	80,61	71,78

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
Episódicos	GB	84,98	89,03	92,74	68,17	80,46	65,22
Relatos	GB	89,37	89,03	92,74	68,17	80,46	65,22
Relatos	1DCNN	92,2	89,04	92,74	68,18	80,46	65,22
Atividade	1DCNN	86,8	85,61	87,9	72,72	80,32	60,38
AER	GB	89,13	91,78	96,77	63,63	80,21	70,0
AER	1DCNN	94,65	91,78	96,77	63,63	80,21	70,0
ARM	1DCNN	93,99	91,78	96,77	63,64	80,21	70
ER	DNN	88,22	78,77	78,23	81,82	80,02	53,73
AERM	GB	89,63	91,10	95,97	63,63	79,80	68,28
AR	GB	90,22	91,10	95,97	63,63	79,80	68,28
ER	ET	96,06	91,10	95,97	63,63	79,80	68,28
AM	1DCNN	91,4	87,67	91,13	68,18	79,66	62,5
AEM	CB	86,59	90,41	95,16	63,63	79,4	66,67
Episódicos	CB	85,85	89,03	93,55	63,63	78,59	63,63
EM	1DCNN	84,1	75,34	75	77,27	78,48	76,14
AE	SVM	73,19	88,36	92,74	63,63	78,19	62,22
ER	GB	89,53	88,36	92,74	63,63	78,19	62,22
ARM	DNN	85,44	84,93	87,9	68,17	78,03	57,69
AERM	ET	96,48	91,10	96,77	59,08	77,92	66,67
AM	GB	84,64	91,10	96,77	59,08	77,92	66,67
AE	1DCNN	88,5	87,67	91,94	63,63	77,79	60,87
AEM	RF	94,15	90,41	95,97	59,08	77,53	65,0
ERM	ET	95,89	90,41	95,97	59,08	77,53	65,0
AEM	SVM	72,88	86,99	91,13	63,63	77,38	59,57
ARM	ET	95,75	89,73	95,16	59,08	77,13	63,41
AEM	1DCNN	89,3	89,73	95,16	59,08	77,13	63,41
RM	1DCNN	93,42	89,73	95,16	59,08	77,13	63,41
ERM	1DCNN	94,06	86,3	90,32	63,63	76,98	58,33
ER	RF	95,46	89,03	94,35	59,08	76,72	61,9
AE	ET	95,44	91,10	97,58	54,55	76,06	64,86
AE	GB	86,77	91,10	97,58	54,55	76,06	64,86
AEM	GB	86,99	91,10	97,58	54,55	76,06	64,86
Mobilidade	RF	91,09	91,10	97,58	54,55	76,06	64,86
AER	ET	95,24	90,41	96,77	54,55	75,66	63,16
AE	RF	93,85	89,73	95,97	54,55	75,26	61,53
EM	RF	94,76	89,73	95,97	54,55	75,26	61,53

Modalidade	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
EM	ET	95,99	89,73	95,97	54,55	75,26	61,53
AE	DNN	84,55	86,3	91,13	59,08	75,11	56,52
RM	DNN	85,38	89,03	95,16	54,55	74,85	60,0
Episódicos	SVM	67,9	85,61	90,32	59,08	74,71	55,32
AEM	DNN	85,3	90,41	97,58	50,0	73,79	61,11
AM	DNN	83,93	86,99	92,74	54,55	73,64	55,81
ARM	SVM	75,75	86,99	92,74	54,55	73,64	55,81
RM	SVM	75,75	86,99	92,74	54,55	73,64	55,81
AM	ET	95,17	89,73	96,77	50,0	73,39	59,46
Episódicos	DNN	81,78	89,03	95,97	50,0	72,98	57,89
Relatos	SVM	75,22	84,93	90,32	54,55	72,43	52,17
Atividade	GB	82,84	87,67	94,35	50,0	72,18	55,0
EM	GB	88,02	87,67	94,35	50,0	72,18	55,00
Mobilidade	1DCNN	84,8	78,08	80,65	63,64	72,14	46,67
AM	RF	93,28	89,73	97,58	45,45	71,52	57,14
AERM	SVM	76,87	86,3	92,74	50,0	71,37	52,38
Mobilidade	GB	84,44	86,3	92,74	50,0	71,37	52,38
AEM	ET	95,61	89,03	96,77	45,45	71,11	55,55
AER	SVM	76,37	84,25	90,32	50,0	70,16	48,89
AM	SVM	67,36	84,25	90,32	50,0	70,16	48,89
EM	DNN	84,23	88,36	96,77	40,91	68,84	51,43
Mobilidade	ET	94,27	88,36	96,77	40,91	68,84	51,43
Atividade	CB	83,43	84,93	91,94	45,45	68,7	47,62
Atividade	DNN	83,28	80,82	86,29	50,0	68,15	44,0
Atividade	RF	92,31	86,99	95,16	40,91	68,04	48,65
Atividade	ET	94,16	86,99	95,16	40,91	68,04	48,65
ERM	SVM	76,32	86,99	95,16	40,91	68,04	48,65
Episódicos	1DCNN	81,2	86,99	95,16	40,91	68,04	48,65
AR	SVM	75,37	84,93	92,74	40,91	66,83	45,0
Mobilidade	SVM	61,3	77,4	82,26	50,0	66,13	40,0
ER	SVM	76,37	83,56	91,13	40,91	66,02	42,86
ER	1DCNN	92,2	87,67	97,58	31,82	64,7	43,75
Atividade	SVM	66,93	76,71	82,26	45,45	63,85	37,04
Mobilidade	DNN	77,53	81,51	90,32	31,81	61,07	34,15
AERM	1DCNN	93,32	85,62	100	4,55	52,27	8,7

C.2 Previsão do humor deprimido futuro

Nós comparamos modelos multimodais (Experimento Multimodal ARM) e multitarefa (Classificação binária e regressão) usando abordagens nomotética e idiográfica para prever o humor deprimido em amostras futuras coletadas por adolescentes.

Na primeira atividade, dividimos os dados dos adolescentes semanalmente, os dados coletados entre os dias D1 e D8 foram usados para o treinamento enquanto os dados coletados entre D9 e D14 para teste. Na segunda atividade, avaliamos a capacidade dos modelos para prever, os sMFQ respondidos pelos adolescentes durante o período de coleta de dados. Os resultados de todos os experimentos desenvolvidos são apresentados nas Seções C.2.1 e C.2.2.

C.2.1 Previsão da próxima semana

C.2.1.1 Previsão dos escores binários do questionário sMFQ

Tabela 22 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**. Previsão dos escores binários do sMFQ usando agregação fixa.

Experimento	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
ARM	CB	82,81	90,48	96,88	47,37	72,12	56,25
	ET	87,20	89,12	99,22	21,05	60,14	33,33
	GB	81,59	88,44	95,31	42,11	68,71	48,48
	RF	87,45	88,44	98,44	21,05	59,75	32,0
	SVM	71,96	89,12	95,31	47,37	71,34	52,94
	DNN	71,20	89,12	94,53	52,63	73,58	55,56
	1DCNN	66,95	91,84	97,66	52,63	75,14	62,50

Tabela 23 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**. Previsão dos escores binários do sMFQ usando agregação cumulativa.

Experimento	Classificador	B.Acc média	Acc	Esp	Sen	B. Acc	F1
ARM	CB	86,98	93,12	96,27	75,00	85,64	76,36
	ET	90,47	93,12	96,27	75,00	85,64	76,36
	GB	80,93	92,59	94,41	82,14	88,28	76,67
	RF	89,86	93,12	96,89	71,43	84,16	75,47
	SVM	75,13	91,01	92,55	82,14	87,34	73,02
	DNN	71,57	93,65	98,14	67,86	83,00	76,00
	1DCNN	74,62	90,48	91,93	82,14	87,03	71,87

C.2.1.2 Previsão da intensidade dos sintomas da depressão com base nos escores do questionário sMFQ

Tabela 24 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**. Previsão da intensidade dos sintomas da depressão usando agregação fixa.

Experimento	Classificador	MAE	RMSE
ARM	CB	3,71	4,96
	ET	3,82	4,92
	GB	4,14	5,35
	RF	4,16	5,28
	SVM	3,89	5,58
	DNN	4,04	5,21
	1DCNN	3,91	5,17

Tabela 25 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**.Previsão da intensidade dos sintomas da depressão usando agregação cumulativa.

Experimento	Classificador	MAE	RMSE
ARM	CB	2,87	4,31
	ET	2,84	3,94
	GB	3,00	4,43
	RF	3,39	4,56
	SVM	3,89	5,42
	DNN	2,72	3,86
	1DCNN	2,69	4,28

C.2.1.3 Tarefa: Previsão dos escores binários e da intensidade dos sintomas da depressão com base no questionário sMFQ

Tabela 26 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**. Experimento multitarefa usando agregação fixa.

Experimento	Classificador	Acc	Esp	Sen	B. Acc	F1	MAE	RMSE
ARM	MT1DCNN	89,12	92,97	63,16	78,06	60,0	4,06	5,29

Tabela 27 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem nomotética**. Experimento multitarefa usando agregação cumulativa.

Experimento	Classificador	Acc	Esp	Sen	B. Acc	F1	MAE	RMSE
ARM	MT1DCNN	95,23	96,27	89,28	92,77	84,74	2,66	3,91

C.2.1.4 Tarefa: Previsão dos escores binários e da intensidade dos sintomas da depressão com base no questionário sMFQ

Tabela 28 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem idiográfica**. Experimento multitarefa usando agregação fixa.

Experimento	Classificador	Acc	Esp	Sen	B. Acc	F1	MAE	RMSE
ARM	MT1DCNN-J	89,79	99,21	26,31	62,67	40,0	3,81	5,02
	MT1DCNN-E	91,83	98,43	47,36	72,90	60,0	3,72	4,92
	MT1DCNN-C	89,79	100	21,05	60,52	34,78	3,80	5,09

Tabela 29 – Humor deprimido futuro: Previsão da próxima semana. **Abordagem idiográfica**. Experimento multitarefa usando agregação cumulativa.

Experimento	Classificador	Acc	Esp	Sen	B. Acc	F1	MAE	RMSE
ARM	MT1DCNN-J	94,70	98,13	75,0	86,56	80,76	2,52	3,53
	MT1DCNN-E	95,23	96,89	85,71	91,30	84,21	2,23	3,58
	MT1DCNN-C	94,70	97,51	78,57	88,04	81,19	2,31	3,71

C.2.2 Previsão dos próximos sMFQ

C.2.2.1 Tarefa: Previsão dos escores binários e da intensidade dos sintomas da depressão com base no questionário sMFQ

Tabela 30 – Humor deprimido futuro: Previsão dos próximos sMFQ. **Abordagem nomotética**. Experimento multitarefa usando agregação fixa.

Experimento	Classificador	sMFQ	Acc	B. Acc	MAE	RMSE
ARM	MT1DCNN	2º	14,29	50,00	4,54	7,09
		3º	69,39	46,43	4,43	6,38
		4º	87,76	72,56	4,83	7,62
		5º	89,80	65,50	4,52	7,46
		6º	79,59	74,03	4,50	7,75
		7º	87,76	75,00	4,51	7,26

Tabela 31 – Humor deprimido futuro: Previsão dos próximos sMFQ. **Abordagem nomotética**. Experimento multitarefa usando agregação cumulativa.

Experimento	Classificador	sMFQ	Acc	B. Acc	MAE	RMSE
ARM	MT1DCNN	2º	84,13	50,00	5,38	8,06
		3º	77,78	82,41	4,96	7,02
		4º	88,89	80,39	5,14	7,84
		5º	82,54	89,81	4,97	8,06
		6º	92,06	81,48	5,10	8,39
		7º	92,06	83,11	5,03	7,99

C.2.2.2 Tarefa: Previsão dos escores binários e da intensidade dos sintomas da depressão com base no questionário sMFQ

Tabela 32 – Humor deprimido futuro: Previsão dos próximos sMFQ. **Abordagem idiográfica**. Experimento multitarefa usando agregação fixa.

Experimento	Classificador	sMFQ	Acc	B. Acc	MAE	RMSE
ARM	MT1DCNN-E	2º	85,71	79,76	3,98	5,05
		3º	84,69	64,28	3,82	4,96
		4º	85,03	64,98	3,73	4,91
		5º	86,22	63,69	3,74	4,96
		6º	87,34	64,28	3,82	5,15
		7º	88,43	66,71	3,66	5,00

Tabela 33 – Humor deprimido futuro: Previsão dos próximos sMFQ. **Abordagem idiográfica**. Experimento multitarefa usando agregação cumulativa.

Experimento	Classificador	sMFQ	Acc	B. Acc	MAE	RMSE
ARM	MT1DCNN-E	2º	84,12	78,39	3,37	4,71
		3º	86,50	76,90	3,09	4,34
		4º	86,24	72,32	2,98	4,17
		5º	89,28	77,40	2,84	3,98
		6º	89,52	77,98	2,85	4,04
		7º	89,94	78,15	2,78	3,97

APÊNDICE D – Modelos de AM unimodais e multimodais desenvolvidos na tese

D.1 Algoritmos de aprendizado clássico

A Tabela 34 apresenta o conjunto de hiperparâmetros dos algoritmos de aprendizado clássico utilizados neste estudo e os hiperparâmetros de configuração, os resultados obtidos na Previsão do humor deprimido atual.

Tabela 34 – Espaço de hiperparâmetros dos algoritmos de aprendizado clássicos.

Hiperparâmetros	Intervalo	RF*	RF**	CB	GB
class_weight	-	balanced			
criterion	-	entropy	entropy	-	friedman_mse
n_estimators	1 - 1000	556	334	778	778
C	1 - 1000			-	
kernel	linear, rbf, poly, sigmoid			-	
gamma	0.001, 0.1, 1, 10, 100			-	
probability	-				
learning_rate	0.0001, 0.001, 0.01, 0.1	-		0.1	0.1

O classificador RF* alcançou o melhor desempenho no experimento unimodal utilizando os áudios episódicos e agregação cumulativa. Ele obteve boas taxas de Acurácia (91,10), Especificidade (95,16), Acurácia Balanceada (81,67). Contudo, os baixos valores de sensibilidade (68,17) e Medida-F1 (69,77) demonstram a dificuldade do modelo para identificar amostras da classe deprimida.

O modelo preditivo formado por classificador CatBoost, agregação cumulativa e os atributos de mobilidade para prever os escores binários de depressão dos questionários sMFQ obteve o desempenho semelhante ao experimento unimodal com áudios episódicos e a mesma dificuldade para prever amostras da classe deprimida.

No experimento unimodal que utilizou os atributos espectrais extraídos dos áudios de relatos, o modelo que obteve o melhor desempenho utilizou o classificador RF**, atingindo taxas de Acurácia de 91,78%, Especificidade de 95,16%, Sensibilidade de 72,72%, Medida-F1 de 72,72%. Além disso, O classificador RF associado a uma agre-

gação cumulativa produziu o melhor modelo unimodal, atingindo 83,94% de Acurácia Balanceada na tarefa de classificação binária da depressão com base na nota de corte do sMFQ.

O experimento Multimodal ARM usou atributos extraídos de dados de Atividade, Áudio de Relatos e Mobilidade. O modelo que alcançou o melhor desempenho utilizou classificador GB e alcançou taxas de 92,47% de Acurácia, 95,16% de Especificidade, 77,27% de Sensibilidade e 75,56% de F1. Consideramos que esse modelo produziu o melhor desempenho entre os experimentos para prever as pontuações binárias do sMFQ, na atividade de previsão do humor atual, alcançando uma taxa de Acurácia Balanceada de 86,22%.

D.2 Algoritmos de aprendizado profundo

A Tabela 35 apresenta os espaços de hiperparâmetros dos algoritmos de aprendizado profundo utilizados neste estudo. Os modelos de aprendizado profundo foram treinados por 100 épocas com uma taxa de aprendizagem inicial de 0.001 que vai sendo ajustada automaticamente durante o treinamento do modelo.

Tabela 35 – Espaço de hiperparâmetros dos algoritmos de aprendizado profundo.

Hiperparâmetros	Intervalo	1DCNN	MT1DCNN*	MT1DCNN**
dense_layers	1-3	3	2	2
dense_layers1	1-3	-	-	-
num_neurons	16, 32, 64, 128	64	128	64
num_neurons1	16, 32, 64, 128	-	-	-
dropout_rate	0 - 0.3	0.2	0.2	0.2
kernel_regularizer	0.1, 0.01, 0.001	0.001	0.1	0.01
cnn_layers	1-2	1	2	2
num_filters	64, 128, 192, 224, 256	64	256	224
kernel_size	1 - 15	15	3	2
pool_size	1 - 4	4	2	2

- *Early Stopping*: interrompe o treinamento antecipadamente, caso não haja melhorias significativas na função de perda após 10 épocas consecutivas.
- *Epochs*: o número de vezes em que todo o conjunto de dados de treinamento é utilizado pelo classificado. O modelo pode ser treinado pelo número máximo de 100 épocas.
- *Optimizer*: O otimizador Adam foi adotado durante o treinamento em todos os experimentos.

- *Batch size*: O tamanho do lote usado durante o treinamento é 32.
- *Activation*: Função de ativação ReLU foi usada nas camadas ocultas das redes neurais.
- *Kernel initializer*: O método de inicialização de pesos foi usado o *glorot uniform*.

O modelo que obteve o melhor desempenho no experimento unimodal usando dados de Atividade utilizou rede convolucional 1DCNN e obteve valores de Acurácia e Acurácia Balanceada de 85,61% e 80,32%, respectivamente. Esse modelo obteve resultados inferiores aos demais experimentos realizados para a previsão do humor deprimido atual.

O modelo MT1DCNN foi desenvolvido especificamente para as atividades de previsão do estado de humor futuro e avaliado em tarefas de classificação binária e regressão. As Tabelas 36 e 37 descrevem os modelos nomotéticos usados nas atividades de previsão da próxima semana e dos próximos sMFQ, respectivamente.

Tabela 36 – Previsão da próxima semana: Parâmetros de configuração do Modelo MT1DCNN

Camadas	**Parâmetros de configuração**
Conv1D	Layers: 2 Filters: 256 Batchnormalization Maxpooling1D Dropout: 0.2
Flatten	
Dense	Layers: 2 Hidden: 128 Ativação: ReLU Kernel regularizer: 0.1 Dropout:0.2
Saída de classificação	Sigmóide
Saída de regressão	Linear

Em síntese, o modelo recebe como entrada os atributos de Atividade, Relatos e Mobilidade (ARM) em um vetor de 239 dimensões. O modelo é formado por dois blocos de convolução *(Conv1D, BatchNormalization, MaxPooling1D e dropout)*. Osblocos têm 256 filtros, um kernel de tamanho 3 e taxa de dropout de 20%. O vetor de saída resultante após a conversão da *flatten* é passado a duas camadas densas com 128 unidades ocultas e função de ativação ReLU, adicionadas à camada de *dropout* com taxa de 20%. Ao final, o modelo tem duas camadas de saída, uma para classificação binária e outra para regressão.

Tabela 37 – Previsão dos próximos sMFQ: Parâmetros de configuração do Modelo MT1DCNN

Camadas	**Parâmetros de configuração**
Conv1D	Layers: 2 Filters: 224 Batchnormalization Maxpooling1D Dropout: 0.2
Flatten	
Dense	Layers: 2 Output_size: 64 Ativação: ReLU Kernel regularizer: 0.01 Dropout:0.2
Saída de classificação	Sigmóide
Saída de regressão	Linear

O conjunto ótimo de hiperparâmetros do modelo MT1DCNN desenvolvido para a previsão dos próximos sMFQ é bastante semelhante ao modelo anterior. Exceto pelos dois blocos de convolução com 224 filtros, duas camadas densas com 64 unidades ocultas com taxa de regularização de kernel de 0.01.

APÊNDICE E – Importância de atributos

Nesta seção apresentamos as matrizes de confusão e os gráficos de importância de atributos dos modelos que obtiveram os melhores desempenhos durante os experimentos para a previsão do humor deprimido atual e futuro.

E.1 Gráficos de importância dos atributos

A matriz de confusão permite visualizar o desempenho dos classificadores adotados neste trabalho. Ela contém os valores reais e preditos pelos modelos de classificação binária de depressão adotados neste estudo.

A seguir, destacamos as principais informações derivadas da matriz de confusão e dos três gráficos de importância de atributos dos modelos que obtiveram o melhor desempenho em cada experimento da tese. São eles: Importância global, direcionalidade e direcionalidade por atributos.

Para a previsão do humor deprimido atual, apresentamos, os gráficos dos melhores modelos nos experimentos unimodais usando dados de Atividade (Figuras 33 a 36); Mobilidade, (Figuras 37 a 40); Áudios episódicos (Figuras 41 a 44); Áudios de Relatos, (Figuras 45 a 48); e do modelo com o melhor desempenho no experimento multimodal (ARM) que combinada dados de Atividade, Áudio de Relatos e Mobilidade (Figuras 21 a 24).

Para a previsão do humor deprimido futuro, apresentamos os gráficos do modelo que obteve o melhor desempenho na atividade de previsão das amostras da próxima semana (Figuras 27 a 30). Cabe ressaltar que não foram gerados os gráficos para os modelos de previsão dos próximos sMFQ devido à alta demanda de processamento exigida.

As matrizes de confusão e os gráficos de SHAP do experimento multimodal ARM para a previsão do humor deprimido atual e futuro são apresentados no texto principal da tese. Essas informações nos possibilitaram identificar associações relevantes entre os modelos preditivos, atributos e os sintomas da depressão.

E.1.1 Matriz de confusão e gráficos de importância dos 20 atributos que mais impactaram a saída do modelo para a previsão do humor deprimido atual. Experimento Unimodal Atividade. Classificador 1DCNN

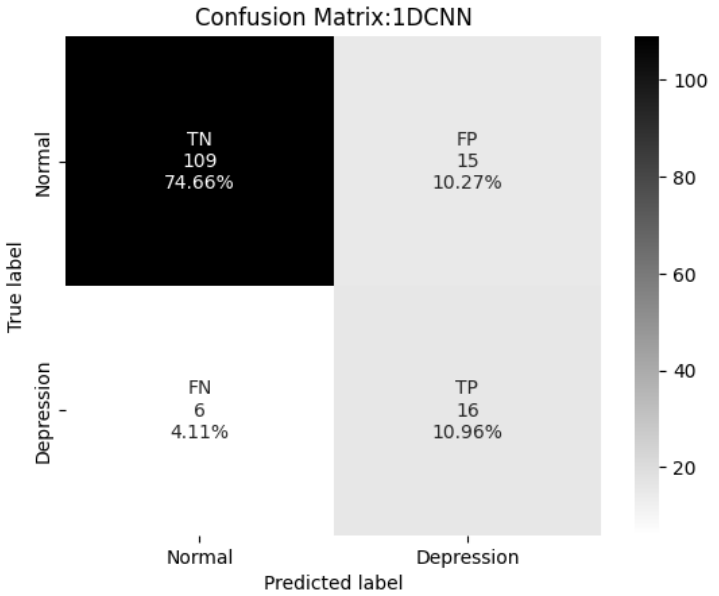


Figura 33 – Experimento Unimodal Atividade: Matriz de confusão

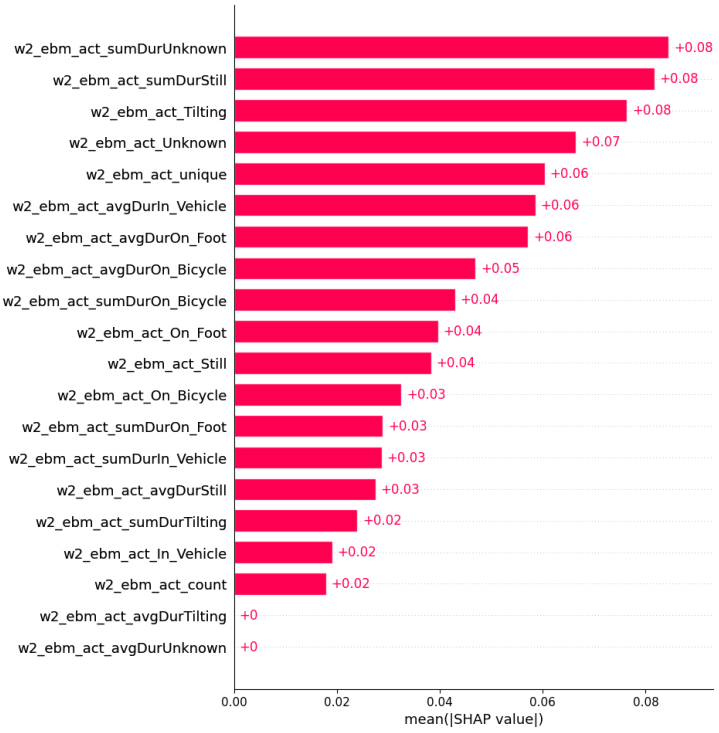


Figura 34 – Experimento Unimodal Atividade: Importância global

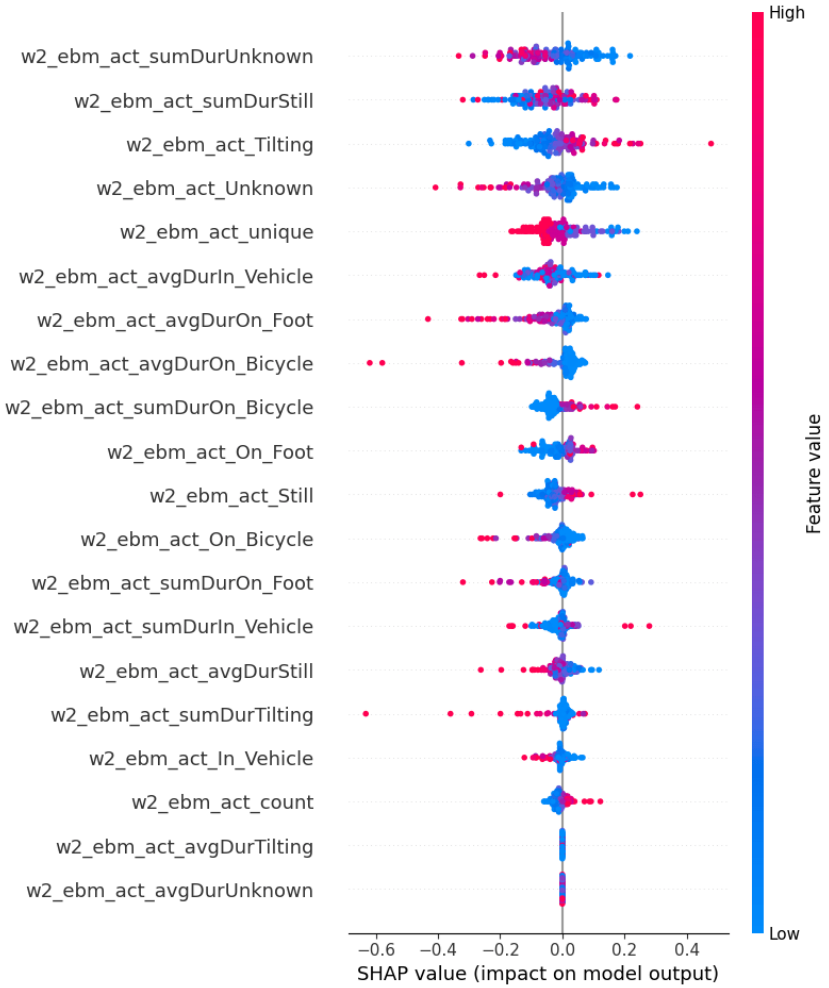


Figura 35 – Experimento Unimodal Atividade: Direcionalidade

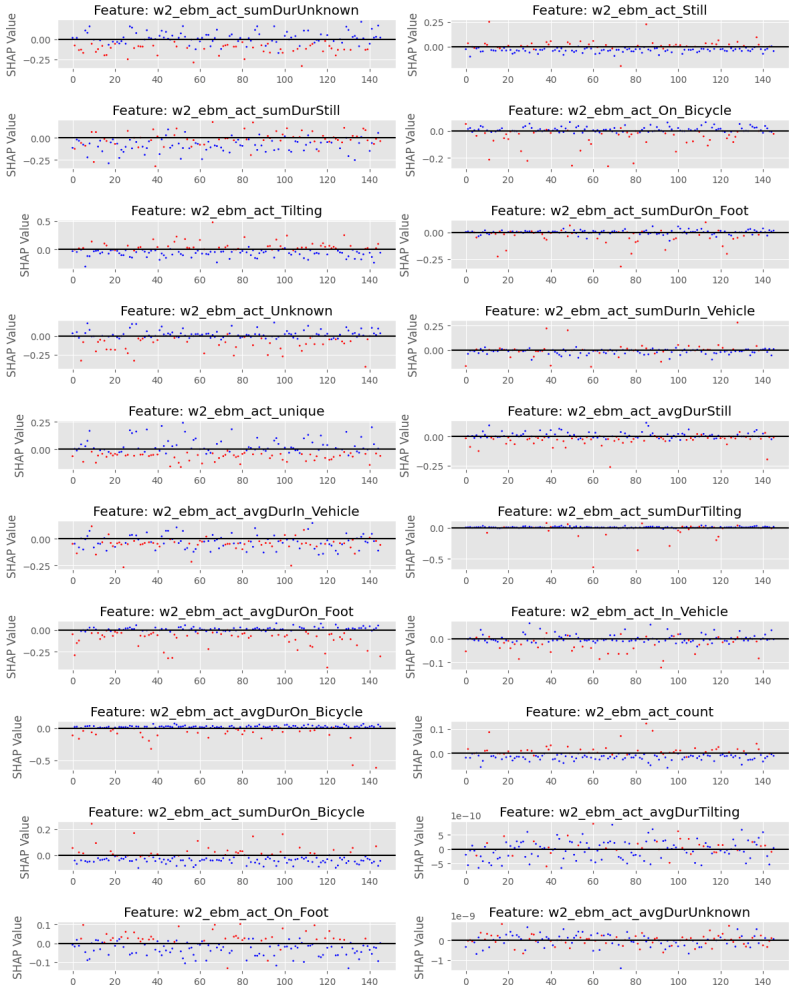


Figura 36 – Experimento Unimodal Atividade: Direcionalidade por atributo

E.1.2 Matriz de confusão e gráficos de importância dos 19 atributos que mais impactaram a saída do modelo para a previsão do humor deprimido atual. Experimento Unimodal Mobilidade. Classificador *CatBoost*

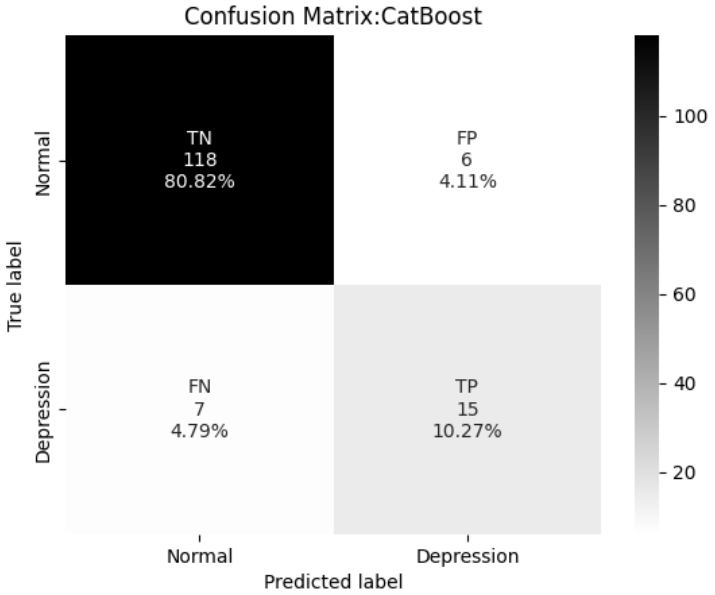


Figura 37 – Experimento Unimodal Mobilidade: Matriz de confusão

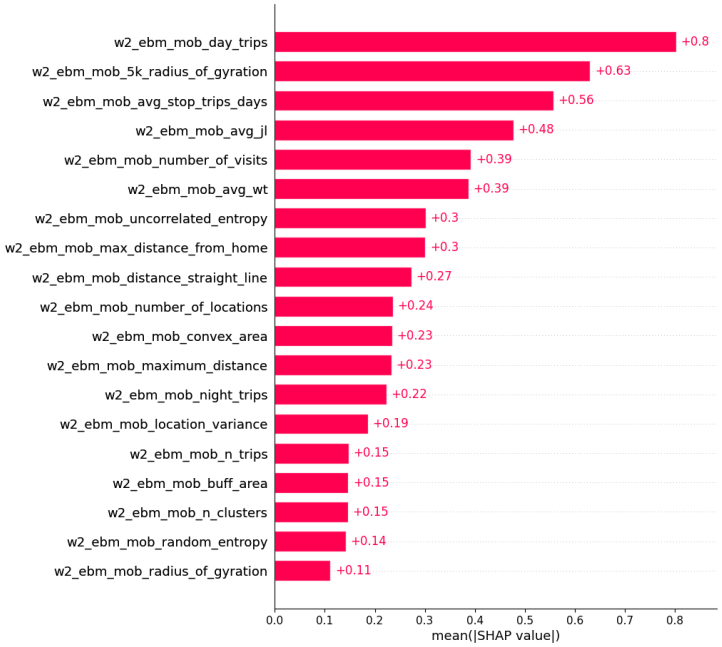


Figura 38 – Experimento Unimodal Mobilidade: Importância global

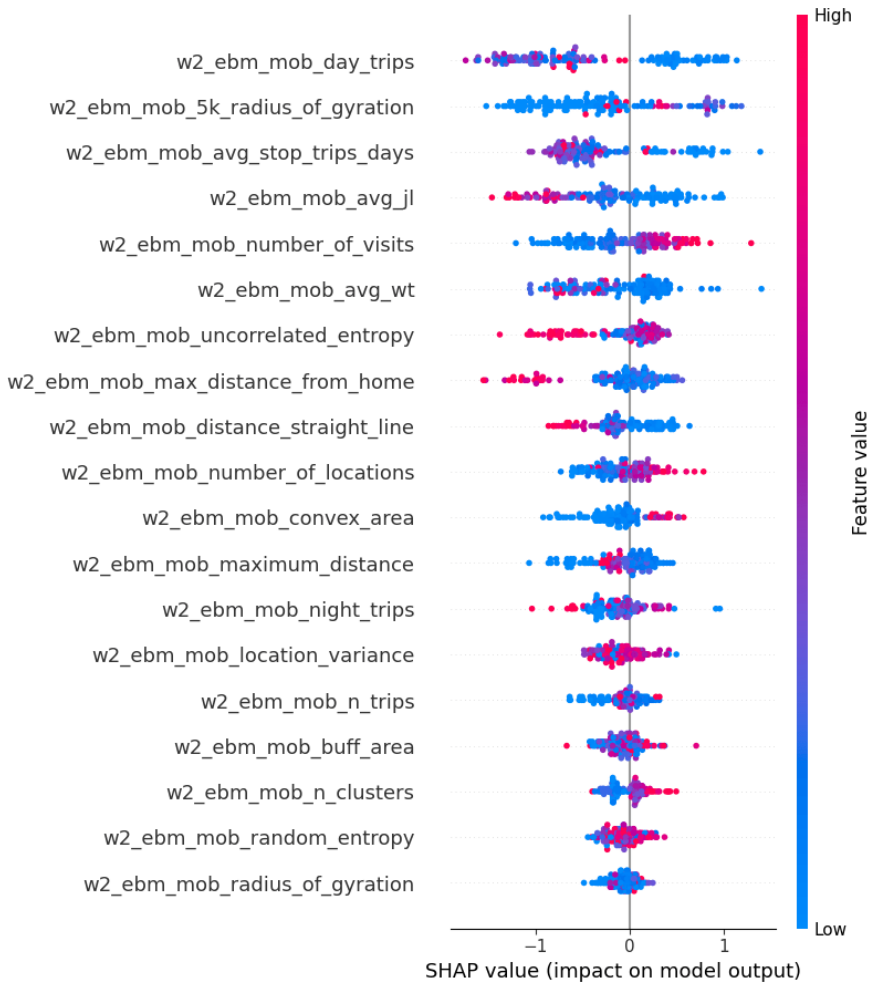


Figura 39 – Experimento Unimodal Mobilidade: Direcionalidade

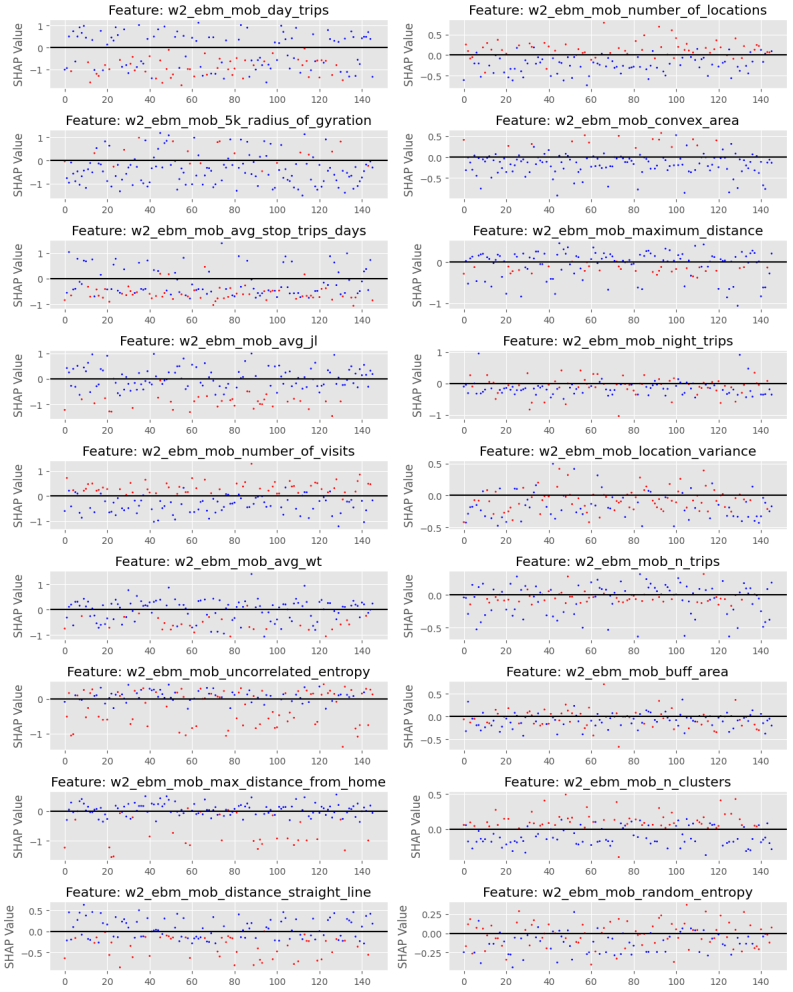


Figura 40 – Experimento Unimodal Mobilidade: Direcionalidade por atributo

E.1.3 Matriz de confusão e gráficos de importância dos 20 atributos que mais impactaram a saída do modelo para a previsão do humor deprimido atual. Experimento Unimodal Áudios Episódicos. Classificador *Random Forest*

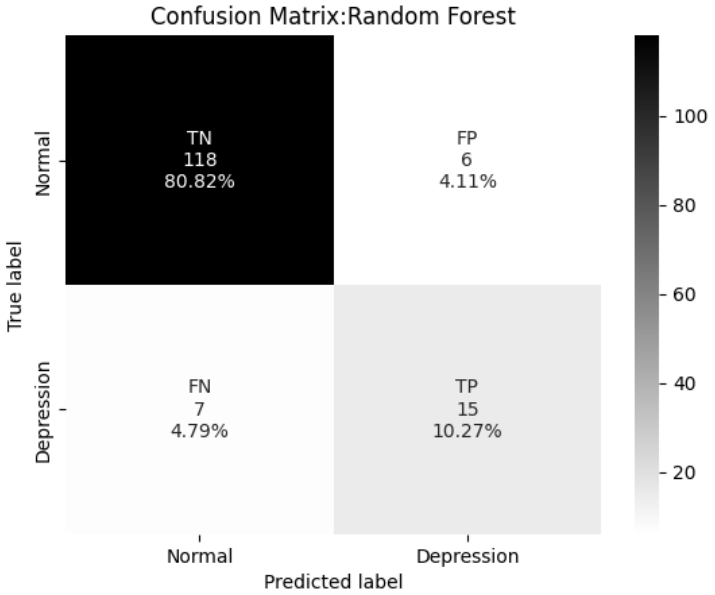


Figura 41 – Experimento Unimodal Áudios Episódicos: Matriz de confusão

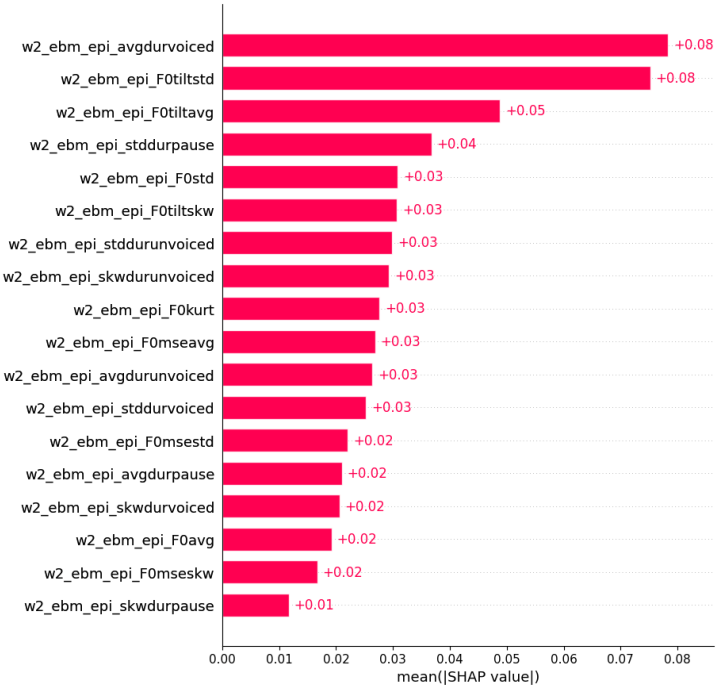


Figura 42 – Experimento Unimodal Áudios Episódicos: Importância global

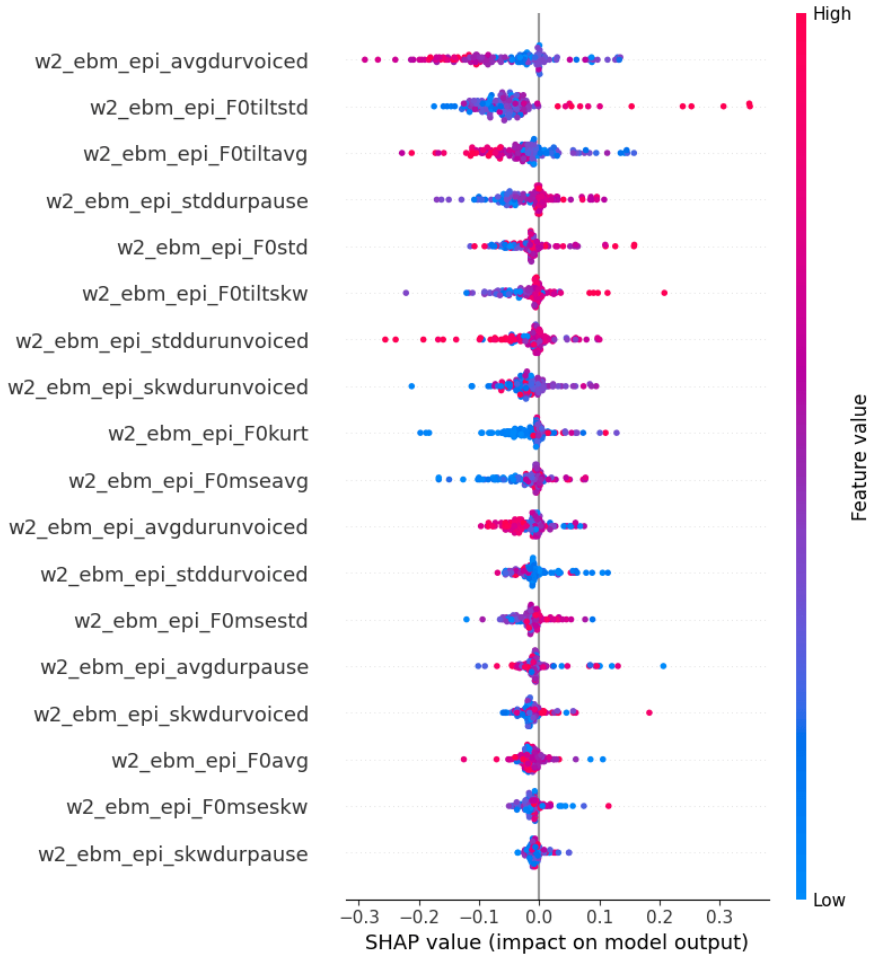


Figura 43 – Experimento Unimodal Áudios Episódicos: Direcionalidade

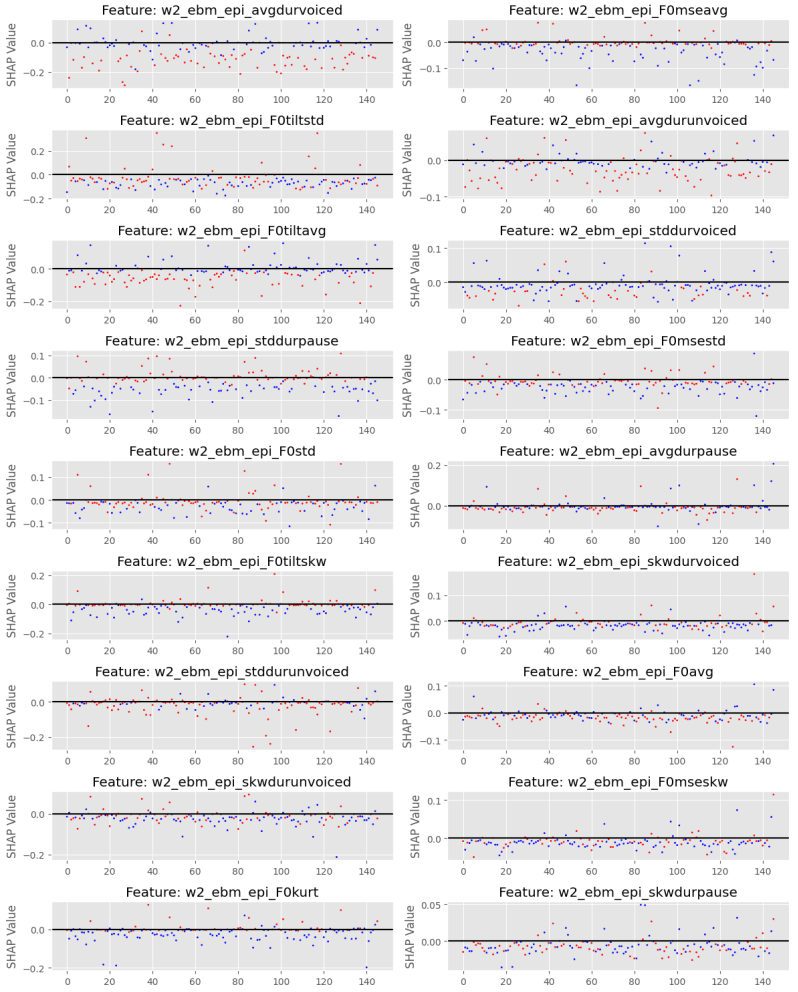


Figura 44 – Experimento Unimodal Áudios Episódicos: Direcionalidade por atributo

E.1.4 Matriz de confusão e gráficos de importância dos 20 atributos que mais impactaram a saída do modelo para a previsão do humor deprimido atual. Experimento Unimodal Áudios de Relatos. Classificador *Random Forest*

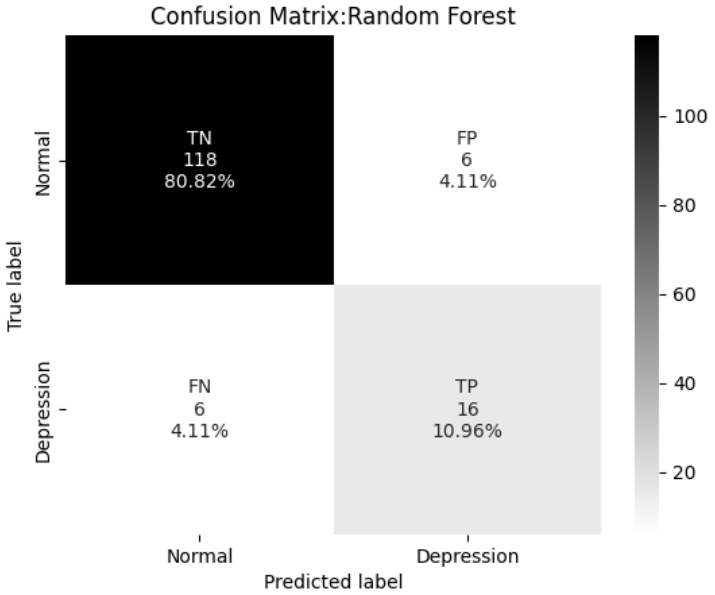


Figura 45 – Experimento Unimodal Áudios de Relatos: Matriz de confusão

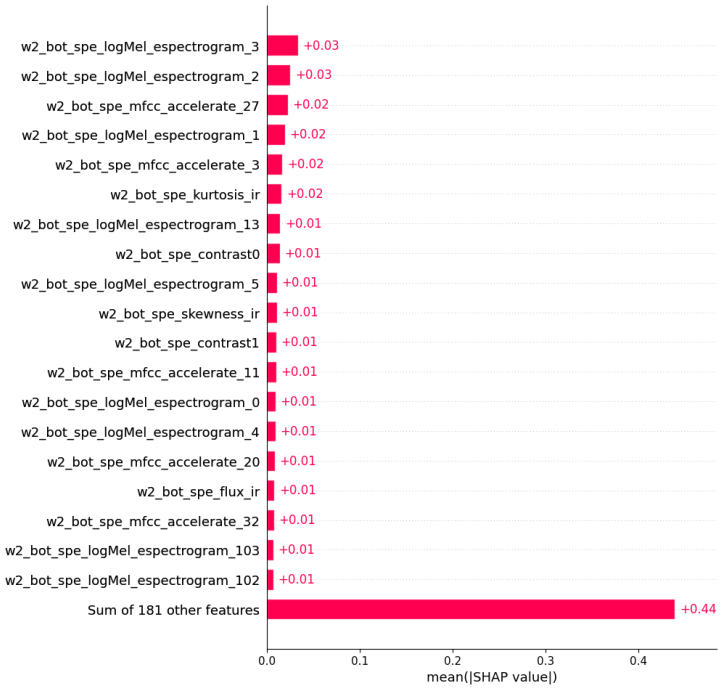


Figura 46 – Experimento Unimodal Áudios de Relatos: Importância global

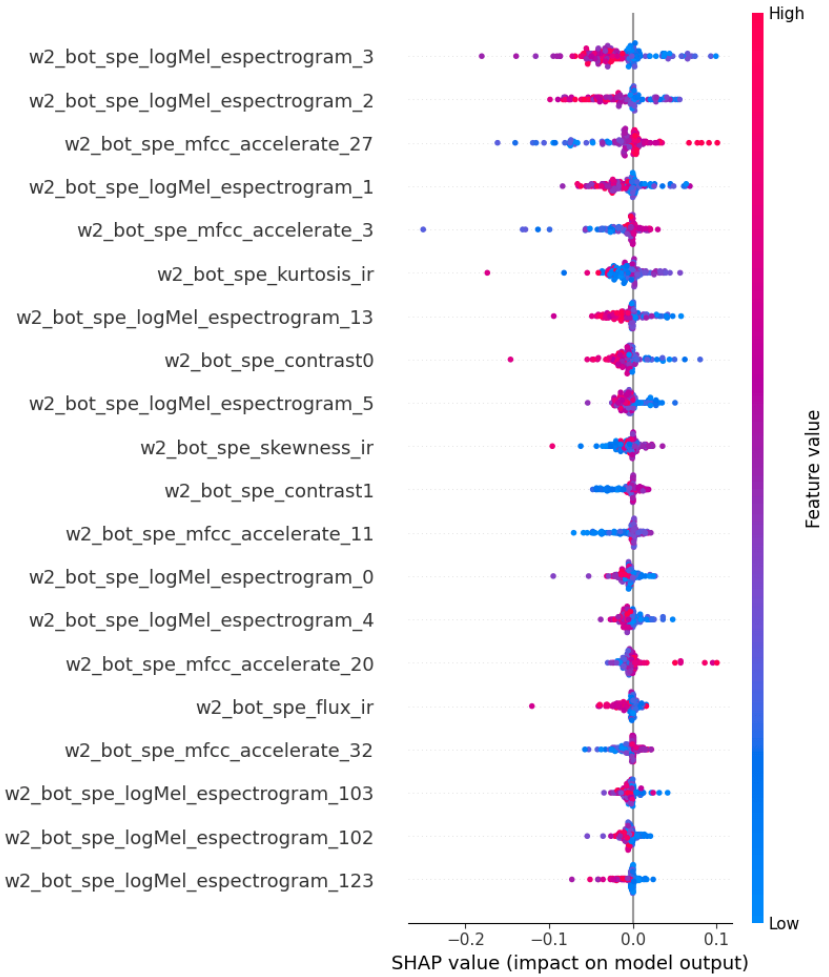


Figura 47 – Experimento Unimodal Áudios de Relatos: Direcionalidade

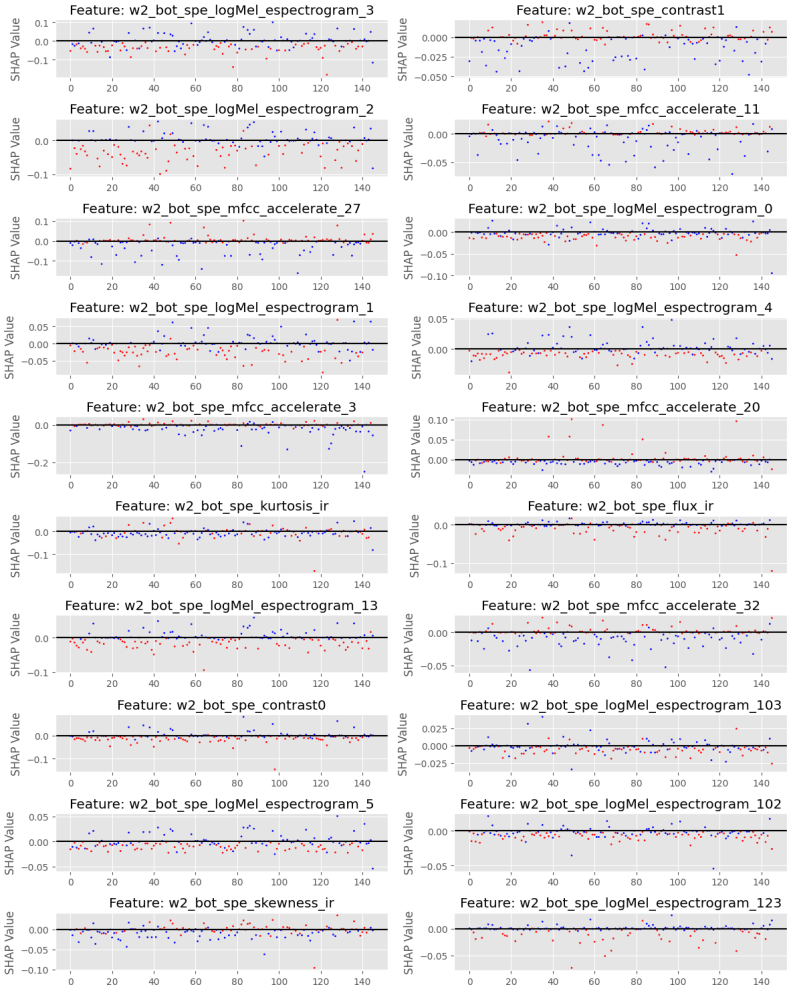


Figura 48 – Experimento Unimodal Áudios de Relatos: Direcionalidade por atributo

APÊNDICE F – Validação do ponto de corte binário do sMFQ

O sMFQ é uma versão reduzida do questionário autorrelatado MFQ, (13 itens) desenvolvido para avaliar a depressão em crianças e adolescentes. Estabelecemos o valor 12 como a nota de corte binária do sMFQ após um processo de validação interna usando as amostras dos adolescentes coletadas no W3.

Nós extraímos as 13 perguntas do MFQ inicial do W3 para gerar um novo smfq simulado. Em seguida, comparamos as pontuações do sMFQ simulado com o escore binário para o diagnóstico de depressão do K-SADS-PL do W3, o questionário padrão-ouro e referência do projeto IDEA-RiSCo.

Esse processo foi repetido para amostras do sexo feminino, do sexo masculino e para todas as amostras usando as medidas Youden e AUC para identificar o limiar capaz de discriminar de maneira mais eficiente amostras deprimidas e não deprimidas do sMFQ.

A Tabela 38 apresenta uma versão do questionário do Humor e Sentimentos - Versão para crianças, traduzido para o português. As pontuações binária e total do sMFQ foram usadas como verdade básica para os modelos preditivos desenvolvidos neste estudo.

Tabela 38 – Questionário do Humor e Sentimentos - Versão para crianças

	Por favor, responda a todas as afirmações abaixo.	Não verdadeiro	Às vezes verdadeiro	Verdadeiro
		0	1	2
1	Eu me senti muito triste ou infeliz.			
2	Eu não consegui me divertir com absolutamente nada.			
3	Eu me senti tão cansado(a) que só ficava sentado(a) sem fazer nada.			
4	Eu estive muito agitado(a).			
5	Eu senti que eu não valia mais nada.			
6	Eu chorei muito.			
7	Eu achei difícil raciocinar ou me concentrar.			
8	Eu me odiei.			
9	Eu me senti uma pessoa ruim.			
10	Eu me senti sozinho(a).			
11	Eu pensei que ninguém me amava de verdade.			
12	Eu pensei que eu nunca seria tão bom(boa) quanto os outros da minha idade.			
13	Eu fiz tudo errado.			